

時空間大規模データに対する統計的解析法

矢島 美寛[†]・平野 敏弘[†]

(受付 2011 年 7 月 1 日；改訂 9 月 1 日；採択 9 月 8 日)

要 旨

本論文では、時空間大規模データを効率的に解析するために提案された統計的解析法の最近の発展についてサーベイする。まず最良不偏線形予測量を近似し、迅速に予測量を計算するために用いられる Covariance Tapering について解説する。次に時空間領域および周波数領域において最尤推定法を代替し、高速計算可能な推定法を紹介する。次にパラメータ数を縮約し少数の因子によりデータを説明する潜在過程モデルを紹介する。最後に今後解決すべき課題を列挙する。

キーワード：時空間統計解析，大規模時空間データ，Covariance Tapering，潜在過程モデル。

1. 序

近年時空間データを収集・整理・解析するためのインフラストラクチャーはリモートセンシング、全地球測位システム (Global Positioning System, GPS)、地理情報システム (Geographical Information System, GIS) などの科学技術の発達、また R や Matlab など統計解析ソフトウェアの普及により整備され充実してきた。

しかしながら時空間データの総数は千、万のオーダーになることは頻繁であり、さらには数十万から百万のオーダーに上ることも稀ではない。したがって大容量の記憶能力および高速の処理能力を保持する現在の計算機をもってしても、このようなサンプル数のデータを迅速に解析することが困難になる場合が生じる。たとえばパラメトリック・モデルの推定や最良線形予測量の構成には、与えられた観測値間の共分散行列の逆行列を計算する必要がある。通常のアプローチでは逆行列の計算に要するオペレーション回数は、サンプル数の 3 乗に比例する。したがってサンプル数が 1,000 個でもオペレーションの回数は 10 億のオーダーになってしまう。

本論文では、21 世紀に入りこの 10 年ほどの間に発展してきた時空間大規模データの効率的な統計的解析法について報告する。

ここで以降の章の構成を述べておく。2 章では準備として時空間データの数学的表現(確率場と言う)、定常確率場およびそれに対する代表的なモデルを説明する。3 章では定常確率場の予測やパラメータ推定に用いられる Covariance Tapering について解説する。4 章では不規則間隔で観測される時空間データに対して、最尤推定法に代わる簡便で効率的な推定法のいくつかを紹介する。5 章ではパラメータ数を極力減らして、データ解析を迅速にする目的で提案された潜在過程モデル (Latent process model) を紹介する。最後に 6 章では今後の課題を列挙する。

なお関連したトピックに関するコンパクトかつ明解なサーベイ論文として Su et al. (2011) が

[†] 東京大学大学院 経済学研究科：〒113-0033 東京都文京区本郷 7-3-1

ある。

2. 時空間データの数学的表現・確率場

本章ではまず時空間データをどのように数学的に表現するかについて説明する。次にこの表現に基づいて定常確率場及びそれに対する代表的なモデルを紹介する。

実数の全体を $\mathbf{R} = (-\infty, \infty)$ とし、その $d(=1, 2, \dots)$ 次元ユークリッド空間は \mathbf{R}^d と表す。また整数の全体を $\mathbf{Z} = \{0, \pm 1, \pm 2, \dots\}$ とし、その d 次元直積集合 $\mathbf{Z} \times \dots \times \mathbf{Z}$ は \mathbf{Z}^d と表す。両者を統一的に表す場合には \mathbf{K}^d とする。

次に観測地点 (site) を $s \in \mathbf{K}^d$ とし、そこで観測されるデータは確率変数 $Y(s)$ と表す。 $Y(s)$ がスカラーのときは一変量データ、ベクトルのときは多変量データである。本論文では一変量データを扱う。例えば $d=2$ のときには、 s は 2 次元ベクトルで第 1, 2 座標は各々緯度、経度を示し、 $Y(s)$ はその地点における地価とする。また $d=3$ のときには、 s は 3 次元ベクトルで、第 1, 2, 3 座標は各々緯度、経度、高さを示し、 $Y(s)$ はその地点における気温などとする。 s の動く領域を $D \subset \mathbf{K}^d$ としたとき、データの全体を $\{Y(s) : s \in D\}$ と書き、確率場 (random field) と言う。

さらに観測時点も考慮する場合には、次元 d を一つ大きくし s の最後の座標が時点を表すとすればよい。先ほどの気温の例では $d=4$ として、 s の第 4 座標を時点に取ればよい。時点を強調したいときには第 4 座標のみ分離して $t \in \mathbf{K}$ と記し、データを $Y(s, t)$ と表す。

確率場 $\{Y(s) : s \in \mathbf{K}^d\}$ が次の 2 条件をみたすとき、弱定常確率場 (weakly stationary random field) と言う。

- (i) 期待値は s に依存せず一定の値 $E(Y(s)) = \mu$ を取る。以下では簡単のため断りのない限り $\mu = 0$ とする。
- (ii) 共分散はベクトル差 $t - s$ のみに依存し、 $C(t - s) = E(Y(t)Y(s))$, $t, s \in \mathbf{K}^d$ となる。

$\{C(\mathbf{h}), \mathbf{h} \in \mathbf{K}^d\}$ を自己共分散関数 (autocovariance function) と言う。以下では弱定常確率場を単に定常確率場と呼ぶ。特に $d=1$ の場合は時系列解析における定常過程になる。 $Y(s)$ および $C(\mathbf{h})$ はスペクトル表現が可能で、それぞれ

$$(2.1) \quad \begin{aligned} Y(s) &= \int_{T^d} \exp(is' \lambda) dM(\lambda) \\ C(\mathbf{h}) &= \int_{T^d} \exp(i\mathbf{h}' \lambda) dF(\lambda) \end{aligned}$$

と表すことが出来る。ここで i は虚数単位 $i^2 = -1$ であり、 $s = (s_1, \dots, s_d)'$, $\mathbf{h} = (h_1, \dots, h_d)'$, $\lambda = (\lambda_1, \dots, \lambda_d)'$ とし、 $'$ は転置ベクトルを意味する。 $\mathbf{K} = \mathbf{Z}$ のときは $T = (-\pi, \pi]$, $\mathbf{K} = \mathbf{R}$ のときは $T = (-\infty, \infty)$ である。 $M = \{M(\lambda), \lambda \in T^d\}$ は d 次元複素直交増分過程、 $F(\lambda)$ は d 次元非負測度でスペクトル分布関数 (spectral distribution function) と各々呼ばれ、任意のボレル集合 $\Delta, \Delta_1, \Delta_2 (\Delta_1 \cap \Delta_2 = \phi) \subset T^d$ に対して、 $E|M(\Delta)|^2 = F(\Delta)$, $E(M(\Delta_1)\overline{M(\Delta_2)}) = 0$ が成立する。スペクトル分布関数が絶対連続な場合には、その密度関数 $f(\lambda)$ をスペクトル密度関数 (spectral density function) と言う (Guyon, 1995; Stein, 1999; Yaglom, 1987)。このとき (2.1) は

$$(2.2) \quad C(\mathbf{h}) = \int_{T^d} \exp(i\mathbf{h}' \lambda) f(\lambda) d\lambda$$

になる。

定常確率場に対して統計モデルを導入する方法は大別して二つに分かれる。ひとつは $Y(s)$ に対して直接導入する方法であり、他方は自己共分散関数 $C(\mathbf{h})$ あるいはスペクトル密度関数

$f(\lambda)$ に導入する方法である。前者にも有用なモデルがあるが、 $d \geq 2$ の場合になると、 $d=1$ の場合すなわち定常時系列のように過去・現在・未来と言う自然な順序がなく、一般にはその意味づけあるいは数学的正当性の証明が難しい。一方自己共分散関数は非負定値関数であればよく、その必要十分条件はスペクトル表現(2.1)を持つことである。またその形状から地点間の観測値の相関関係も把握しやすい。したがってここでは後者のアプローチを採用する。

現時点までに提案されているモデルの中で代表的なものとして等方型モデル (Isotropic Model) と分離型あるいは乗法型モデル (Separable Model) がある。

等方型モデルは、自己共分散関数 $C(\mathbf{h})$ が距離 $\|\mathbf{h}\| = \sqrt{\sum_{i=1}^d h_i^2}$ のみに依存し、方向には無関係なモデルである。すなわち $C_0(x)(x \in \mathbf{R})$ を 1 変数正定値関数としたとき、 $C(\mathbf{h})$ は $x = \|\mathbf{h}\|$ を代入して

$$C(\mathbf{h}) = C_0(\|\mathbf{h}\|)$$

によって表現される。

$C_0(x)$ に対するパラメトリック・モデルのなかで、応用上頻繁に用いられるモデルは Matérn 族 (Matérn, 1947, 1960, 1986; Cressie, 1993; Stein, 1999) である。Matérn 族に関する歴史については Guttorp and Gneiting (2006) を参照されたい。Matérn 族の一般形は

$$C_0(x) = \frac{\pi^{1/2} \phi}{2^{\nu-1} \Gamma(\nu + 1/2) \alpha^{2\nu}} (\alpha|x|)^{\nu} \mathcal{K}_{\nu}(\alpha|x|)$$

によって表現される。ここで α, ν, ϕ は正の定数、また \mathcal{K}_{ν} は変形された Bessel 関数 (modified Bessel function) と呼ばれる特殊関数である。 ϕ が共分散の大きさ、 α が形状、 ν が滑らかさおよび $x \rightarrow \infty$ のときの 0 への収束速度を規定するパラメータである。例えば $\nu = 1/2$ のときは $\mathcal{K}_{1/2}(x) = \sqrt{\pi/(2x)} \exp(-x)$ 、 $\nu = 3/2$ のときは $\mathcal{K}_{3/2}(x) = \sqrt{\pi/(2x)}(1+x^{-1}) \exp(-x)$ になる。したがって自己共分散関数は $\nu = 1/2$ のときは $C_0(x) = \pi\phi\alpha^{-1} \exp(-\alpha|x|)$ 、 $\nu = 3/2$ のときは $C_0(x) = \frac{1}{2}\pi\phi\alpha^{-3} \exp(-\alpha|x|)(1+\alpha|x|)$ になる。この自己共分散関数に対応するスペクトル密度関数は $\omega = \|\lambda\|$ のみに依存し

$$f(\omega) = \frac{\phi}{\pi^{(d-1)/2} (\alpha^2 + \omega^2)^{\nu+d/2}}$$

によって表現される。

次に分離型モデルでは、 $C(\mathbf{h})$ が 2 個以上の正定値関数の積として表現される。例えば \mathbf{h} を 2 つのベクトル $\tilde{\mathbf{h}}_1 = (h_1, \dots, h_m)'$ 、 $\tilde{\mathbf{h}}_2 = (h_{m+1}, \dots, h_d)'$ に分割し、 $C(\mathbf{h})$ は 2 つの正定値関数 $C_1(\tilde{\mathbf{h}}_1)$ 、 $C_2(\tilde{\mathbf{h}}_2)$ の積 $C(\mathbf{h}) = C_1(\tilde{\mathbf{h}}_1)C_2(\tilde{\mathbf{h}}_2)$ によって定義される。

この統計モデルを周波数領域から見ると、 $\lambda = (\lambda_1, \dots, \lambda_d)'$ を 2 つのベクトル $\tilde{\lambda}_1 = (\lambda_1, \dots, \lambda_m)'$ 、 $\tilde{\lambda}_2 = (\lambda_{m+1}, \dots, \lambda_d)'$ に分割し、スペクトル密度関数 $f(\lambda)$ は 2 つの正の値を取る関数 $f_1(\tilde{\lambda}_1)$ 、 $f_2(\tilde{\lambda}_2)$ の積 $f(\lambda) = f_1(\tilde{\lambda}_1)f_2(\tilde{\lambda}_2)$ によって表現される。

ただし等方型あるいは分離型モデルが適さない現実のデータは多々ある。現在非等方型・非分離型のモデルの開発も精力的に行われていることを断っておく。関心のある読者は矢島 (2011) およびそれに所収の参考文献を参照されたい。

3. Covariance Tapering の予測への応用

Covariance Tapering とは、真の自己共分散関数をサポートが有限な自己共分散関数で代替し、さまざまな統計的推測を行う方法である。時系列解析とのアナロジーで言えば、真のモデルをある一定の時間間隔以上離れた観測値が互いに無相関になる移動平均モデルで近似して統計的解析を行うことに対応する。本章では Covariance Tapering の応用例として既知の観測値

から未知の値を予測する問題を考える. なお推定への応用については紙幅の関係で割愛するが, Du et al. (2009), Kaufman et al. (2008), Wang and Loh (2011)などの文献を参照されたい.

まず最良線形不偏予測量(Best Linear Unbiased Predictor, BLUP)を定式化する. 次に特定化の誤りによりあるいは意図的に, 真の自己共分散関数とは異なった自己共分散関数を用いて構成されたBLUPの最小2乗誤差と真のBLUPのそれが漸近的に等しくなるための十分条件を説明する. 次にCovariance Taperingについて説明し, この方法に基づくBLUPが十分条件を満たす場合を明らかにする.

いま回帰モデル

$$Y(\mathbf{s}) = m(\mathbf{s})'\beta + \epsilon(\mathbf{s})$$

を考える. ここで $m(\mathbf{s})$, β は p 次元ベクトルで各々説明変数, 回帰係数ベクトルとし, $\{\epsilon(\mathbf{s})\}$ は期待値0の定常確率場とする. このとき $\mathbf{Y} = (Y(\mathbf{s}_1), Y(\mathbf{s}_2), \dots, Y(\mathbf{s}_n))'$ が観測され, 別の地点 \mathbf{s}^* の値 $Y(\mathbf{s}^*)$ を予測するとしよう. 観測値に対して線形 $\lambda'\mathbf{Y}$ ($\lambda \in \mathbf{R}^n$)でありかつ不偏性 $E(\lambda'\mathbf{Y}) = \lambda'M\beta = m(\mathbf{s}^*)'\beta$ をみたす予測量の中で平均2乗誤差 $E[Y(\mathbf{s}^*) - \lambda'\mathbf{Y}]^2$ を最小にする予測量をBLUPと言う. ここで M は $n \times p$ 行列 $M = (m(\mathbf{s}_1), \dots, m(\mathbf{s}_n))'$ とする.

いま $n \times n$ 行列 C , n 次元ベクトル c^* を

$$C = (c_{ij}), \quad c_{ij} = \text{Cov}(\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j))$$

$$c^* = (c_1^*, \dots, c_n^*)', \quad c_i^* = \text{Cov}(\epsilon(\mathbf{s}^*), \epsilon(\mathbf{s}_i))$$

によって定義する. C は正則行列, M はフル・ランク行列すなわち $\text{rank}(M) = p$ のとき, BLUPおよびその平均2乗誤差は

$$(3.1) \quad \hat{Y}(\mathbf{s}^*) = (c^*)'C^{-1}(\mathbf{Y} - M\hat{\beta}) + m(\mathbf{s}^*)'\hat{\beta}$$

$$(3.2) \quad E[Y(\mathbf{s}^*) - \hat{Y}(\mathbf{s}^*)]^2 = c_0 - (c^*)'C^{-1}c^* + \gamma'M'C^{-1}M\gamma$$

によって与えられる. ここで $\hat{\beta}$ は回帰係数 β の最良線形不偏推定量(Best Linear Unbiased Estimator, BLUE) $\hat{\beta} = (M'C^{-1}M)^{-1}M'C^{-1}\mathbf{Y}$ であり, $\gamma = m(\mathbf{s}^*) - M'C^{-1}c^*$, $c_0 = \text{Var}(\epsilon(\mathbf{s}^*))$ とする(Cressie, 1993).

BLUPは地球統計学の分野ではクリギング(Kriging)とも呼ばれている. クリギングは説明変数を3種類に分け, $p=0$, $m(\mathbf{s}) \equiv 0$ (期待値が既知の場合と同等)を単純クリギング(Simple Kriging), $p=1$, $m(\mathbf{x}) = \text{定数(未知)}$ を通常クリギング(Ordinary Kriging), $m(\mathbf{x})$ が一般の説明変数, β が未知の場合を一般クリギング(Universal Kriging)と言う.

以下では単純クリギング $Y(\mathbf{s}) = \epsilon(\mathbf{s})$ を考える. したがって(3.1)と(3.2)は各々

$$(3.3) \quad \hat{Y}(\mathbf{s}^*) = (c^*)'C^{-1}\mathbf{Y}$$

$$(3.4) \quad E[Y(\mathbf{s}^*) - \hat{Y}(\mathbf{s}^*)]^2 = c_0 - (c^*)'C^{-1}c^*$$

になる.

次に $C_0(\mathbf{h})$ を真の自己共分散関数, $C_1(\mathbf{h})$ を我々が予測量を構成する際に想定する自己共分散関数とする. また $f_i(\lambda)$ ($i=0, 1$)を各々のスペクトル密度関数とする. 真のBLUPと想定された自己共分散関数に基づくBLUPは(3.3)より

$$\hat{Y}(\mathbf{x}^*, C_i) = (c_i^*)'C_i^{-1}\mathbf{Y} \quad (i=0, 1)$$

となる.

サンプル数 n が増加するとき二つの予測量に関して

$$(3.5) \quad \lim_{n \rightarrow \infty} \frac{E_{C_0}[Y(\mathbf{s}^*) - \hat{Y}(\mathbf{s}^*, C_1)]^2}{E_{C_0}[Y(\mathbf{s}^*) - \hat{Y}(\mathbf{s}^*, C_0)]^2} = 1$$

が成立する条件を考えたい。 E_{C_0} は真の自己共分散関数の下での期待値を表す。 予測量の定義から任意の n に対して (3.5) の比は 1 以上であるが、 サンプル数が増加するに連れて想定された自己共分散関数に基づく BLUP の予測誤差が真の BLUP のそれと差がなくなることを等号は意味している。

以下では $\{Y(\mathbf{s})\}$ は正規確率場に従うと仮定する。 ただし予測誤差 (3.4) は確率場の自己共分散関数のみで決まり、 分布に関する他の条件は一切使用しないので、 この仮定は一般性を失わない。 もし $C_i(\mathbf{h})(i=0,1)$ から導入される正規確率場の分布 $P_i(i=0,1)$ が互いに「近ければ」、 関係式 (3.5) の成立が期待できる。 確率分布の「近さ」を測るひとつの概念が同値性 (equivalence) である。

定義 1. Ω を確率空間、 \mathcal{F} を Ω 上で定義された σ -field、 $P_i(i=0,1)$ を確率分布とする。

$$P_0(A) = 0 \iff P_1(A), A \in \mathcal{F}.$$

が成立するとき、 P_0 と P_1 は同値であると言う。

このとき以下の定理が成立する。

定理 1. (Stein, 1990) 確率分布 P_0 と P_1 が同値なとき、 (3.5) が成立する。

定理 1 は非定常確率場においても成立する。 定常確率場に対しては確率分布の同値性が成立するための十分条件がスペクトル密度関数を用いて表現でき、 以下の定理を得る。

定理 2. (Stein, 1993) 以下の条件を仮定する。

$$(3.6) \quad 0 < \liminf_{\|\lambda\| \rightarrow \infty} f_0(\lambda)/|\phi(\lambda)|^2 \leq \limsup_{\|\lambda\| \rightarrow \infty} f_0(\lambda)/|\phi(\lambda)|^2 < \infty$$

$$(3.7) \quad \lim_{\|\lambda\| \rightarrow \infty} f_1(\lambda)/f_0(\lambda) = a (> 0)$$

ここで $\phi(\lambda)$ は band-limited 関数である。 band-limited 関数とはそのフーリエ変換のサポートがコンパクト集合となる関数を言う。 このとき地点 \mathbf{s}^* が $\{\mathbf{s}_i | i=1,2,\dots\}$ の集積点であれば (3.5) が成立する。

ここで定理 2 について補足しておく。 (3.6) はあくまで技術的な仮定であり、 これなしに定理 2 が成立するか否かは未解決な問題である (Stein, 1999)。 また (3.7) は、 値を予測したい地点 \mathbf{s}^* が $\{\mathbf{s}_i | i=1,2,\dots\}$ の集積点のとき、 予測量の精度は想定されたスペクトル密度関数の低周波の挙動には無関係で、 高周波の挙動のみに依存することを意味している。

ここまでの議論においてはサンプル数 n に関わらず我々が想定する自己共分散関数 $C_1(\mathbf{h})$ は固定されていた。 実際のデータ解析においてはサンプル数に応じて $C_1(\mathbf{h})$ を変化させていく方が自然である。 この場合確率分布の近さを意味する概念として、 近接性 (contiguity) がある。

定義 2. $\{P_{0,n}\}, \{P_{1,n}\}$ を確率空間の列 (Ω_n, \mathcal{A}_n) において定義された確率分布の列とする。 任意の $A_n \in \mathcal{A}_n$ に対して、 $P_{0,n}(A_n) \rightarrow 0$ のとき $P_{1,n}(A_n) \rightarrow 0$ が成立するならば、 $\{P_{1,n}\}$ は $\{P_{0,n}\}$ に対して近接していると言う。 また逆も同時に成立するとき、 $\{P_{0,n}\}$ と $\{P_{1,n}\}$ は互いに近接していると言う。

このとき定理 2 は以下のように拡張できる。

定理 3. (Putter and Young, 2001) $n \rightarrow \infty$ のとき $E_{C_0}[Y(\mathbf{s}^*) - \hat{Y}(\mathbf{x}^*, C_0)]^2 \rightarrow 0$ が成立し、 また確率分布列 $\{P_0\}$ と $\{P_{1,n}\}$ は互いに近接しているとする。

このとき $C_{1,n}(\mathbf{h})$ を $P_{1,n}$ によって規定される自己共分散関数とすれば, P_1 を $P_{1,n}$ に置き換えても定理 1 は成立する.

P_0 および $C_0(\mathbf{h})$ は各々真の確率分布, 自己共分散関数であるから n には依存しないことを注意しておく.

定常確率場の場合には, 近接性をスペクトル密度関数で表現すると以下ようになる.

定理 4. (Putter and Young, 2001) $f_{1,n}(\boldsymbol{\lambda})$ を自己共分散関数 $C_{1,n}(\mathbf{h})$ のスペクトル密度関数とする.

このとき $\{P_{1,n}\}$ が $\{P_0\}$ に対して近接するための十分条件は

$$\limsup_{n \rightarrow \infty} n \sup_{\boldsymbol{\lambda}} \left(\frac{f_{1,n}(\boldsymbol{\lambda}) - f_0(\boldsymbol{\lambda})}{f_0(\boldsymbol{\lambda})} \right)^2 < \infty$$

$$\liminf_{n \rightarrow \infty} \inf_{\boldsymbol{\lambda}} \frac{f_{1,n}(\boldsymbol{\lambda})}{f_0(\boldsymbol{\lambda})} < \infty$$

である.

逆に $\{P_0\}$ が $\{P_{1,n}\}$ に対して近接するための十分条件は上式において $f_0(\boldsymbol{\lambda})$ と $f_{1,n}(\boldsymbol{\lambda})$ を入れ替えた条件になる.

Covariance Tapering は真の自己共分散関数をサポートがコンパクト集合になる自己共分散関数によって代替し統計的解析を行う手法である. 上述の予測問題に対する応用方法は以下のようなになる.

いま $C_\theta(\mathbf{h})$ をサポートがコンパクト集合 ($C_\theta(\mathbf{h}) = 0, \|\mathbf{h}\| > \theta$) である自己相関関数 ($C_\theta(\mathbf{0}) = 1$) とする. 真の自己共分散関数 $C_0(\mathbf{h})$ に $C_\theta(\mathbf{h})$ をかけて

$$C_{tap}(\mathbf{h}) = C_0(\mathbf{h})C_\theta(\mathbf{h})$$

を定義する. 図 1 に例を示す. ここでは $C_0(\mathbf{h})$, $C_\theta(\mathbf{h})$ ともに等方型を仮定し, $\theta = 0.8$ としている. $C_{tap}(\mathbf{h})$ が自己共分散関数のときの共分散行列を $\Sigma_{tap} = (\sigma_{ij,tap})(n \times n)$ とすれば,

$$\sigma_{ij,tap} = \begin{cases} C_{tap}(\mathbf{s}_i - \mathbf{s}_j), & \|\mathbf{s}_i - \mathbf{s}_j\| \leq \theta \\ 0, & \|\mathbf{s}_i - \mathbf{s}_j\| > \theta \end{cases}$$

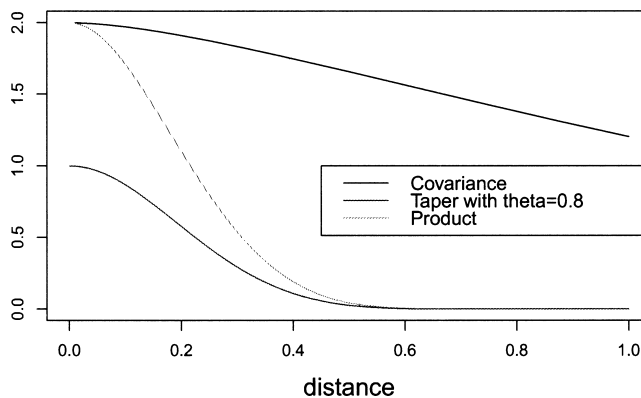


図 1. Covariance Tapering.

となる。したがって Σ_{tap} の成分の多くは 0 となる “sparse” な行列であり、BLUP を構成する際に必要となる逆行列にたいしては高速な計算手法が数値解析の分野において開発されている (Davis, 2006)。

いま $\{Y(\mathbf{s})\}$ は $C_0(\mathbf{h})$ が Matérn 族に属する定常確率場とし、我々が予測に用いる自己共分散関数は $C_1(\mathbf{h}) = C_{tap}(\mathbf{h})$ とする。

このとき以下の定理を得る。

定理 5. (Furrer et al., 2006) $C_\theta(\mathbf{h})$ を等方型の自己相関関数 $C_\theta(\mathbf{h}) = C_\theta(\|\mathbf{h}\|)$ とし、 $f_\theta(\lambda)$ をそのスペクトル密度関数とする。またある $\epsilon(\geq 0)$ と θ のみに依存する定数 $M(\theta) < \infty$ が存在し

$$0 < f_\theta(\boldsymbol{\lambda}) \leq \frac{M(\theta)}{(1 + \|\boldsymbol{\lambda}\|^2)^{\nu+d/2+\epsilon}}$$

をみたすと仮定する。このとき (3.5) が成立する。

Matérn 族に属する自己共分散関数は自動的に定理 2 の仮定 (3.6) をみたすことが知られている (Stein, 1999)。したがって (3.7) が成立することを示せばよい。

4. 最尤推定法に対する近似方法

大規模かつ不規則間隔で観測されるデータに対して最尤推定量を適用するのは、等間隔に観測されるデータにも増して共分散行列の逆行列や行列式の計算が複雑になり実用的とは言えない。より実用的で高速計算可能な近似方法が時空間領域および周波数領域において提案されている。本章ではこれらの方法のいくつかを紹介し、その優劣について議論する。

まず時空間領域において Vecchia (1988) によって提案され、後に Stein et al. (2004) によって一般化された方法を説明する。観測値 $Y(\mathbf{s}_i) (i=1, \dots, n)$ が与えられたとき、 $\mathbf{Y}_n = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))'$ と置く。 $\{Y(\mathbf{s})\}$ が期待値 0 の正規定常確率場にしたがう場合、定数項を除去した θ の対数尤度関数は

$$L_n(\theta; \mathbf{Y}_n) = -\frac{1}{2} \log \det(\Sigma_n) - \frac{1}{2} \mathbf{Y}_n' \Sigma_n^{-1} \mathbf{Y}_n$$

になる。ここで Σ_n は $n \times n$ 共分散行列で、その (i, j) 成分は $\sigma(\mathbf{s}_i, \mathbf{s}_j; \theta) = \text{Cov}(Y(\mathbf{s}_i), Y(\mathbf{s}_j))$ とし、 θ はパラメータ・ベクトルとする。

上述のように行列式 $\det(\Sigma_n)$ や逆行列 Σ_n^{-1} の計算は複雑かつ time-consuming になる。そこでまず \mathbf{Y}_n を b 個の部分ベクトル $\mathbf{Y}_{j_n} (j=1, \dots, b)$ に分割し、 $\mathbf{Y}_n = (\mathbf{Y}'_{1_n}, \dots, \mathbf{Y}'_{b_n})'$ と表す。また $\mathbf{Y}_{(j_n)} = (\mathbf{Y}'_{1_n}, \dots, \mathbf{Y}'_{j_n})'$ と置く。このとき $L_n(\theta; \mathbf{Y}_n)$ は条件付き対数尤度関数の和

$$L_n(\theta; \mathbf{Y}_n) = L_n(\theta; \mathbf{Y}_{1_n}) \sum_{j=2}^b L_n(\mathbf{Y}_{j_n} | \theta; \mathbf{Y}_{(j-1, n)})$$

によって表現できる。ここで $L_n(\mathbf{Y}_{j_n} | \theta; \mathbf{Y}_{(j-1, n)})$ は $\mathbf{Y}_{(j-1, n)}$ が与えられたときの \mathbf{Y}_{j_n} の条件付き対数尤度関数である。

さらに $\mathbf{S}_{(j_n)}$ を $\mathbf{Y}_{(j_n)}$ の部分ベクトルとして、

$$L_n(\theta; \mathbf{Y}_n) \approx L_n(\theta; \mathbf{Y}_{1_n}) \sum_{j=2}^b L_n(\mathbf{Y}_{j_n} | \theta; \mathbf{S}_{(j-1, n)})$$

により対数尤度関数を近似し、 $L_n(\theta; \mathbf{Y}_{1_n}) \sum_{j=2}^b L_n(\mathbf{Y}_{j_n} | \theta; \mathbf{S}_{(j-1, n)})$ を最大にする θ を推定する。真の最尤推定量より計算は簡便になるがその近似精度は \mathbf{Y}_n の分割方法、 $\mathbf{S}_{(j_n)}$ の選択などに依存する。

周波数領域における推定方法として2つ紹介する. ひとつは Fuentes (2007) により提案された方法である. 簡単のため $d=2$ とする. まず $t=(t_1, t_2)'$ を中心とする一辺の長さが Δ の正方形上で $Y(s)$ を積分して, $\{X(t)\}$ を

$$(4.1) \quad X(t) = \Delta^{-2} \int \int_{|t_i - s_i| \leq \Delta/2 (i=1,2)} Y(s) ds$$

によって定義する. このとき $\{X(t)\}$ も定常確率場になり, そのスペクトル密度関数は

$$f_X(\lambda) = \Delta^{-4} |\Gamma(\lambda)|^2 f_Y(\lambda)$$

である. ここで $\Gamma(\lambda) = \prod_{i=1}^2 2 \sin(\Delta \lambda_i / 2) / \lambda_i$ とする.

次に $\{X(t)\}$ を離散地点 $\Delta z (z \in \mathbf{Z}^2)$ においてサンプリングした定常確率場 $\{X(\Delta z)\}$ を考える. $\{X(\Delta z)\}$ のスペクトル密度関数は

$$(4.2) \quad f_{\Delta X}(\lambda) = \sum_{Q \in \mathbf{Z}^2} f_X(\lambda + 2\pi Q / \Delta), \quad \lambda \in [-\pi / \Delta, \pi / \Delta]^2.$$

となる. $\{X(\Delta z)\}$ に対して時系列解析における最尤推定法を近似する方法として用いられる Whittle 推定法を適用する. いま $\Delta \mathbf{X}_m = (X(\Delta z_1), \dots, X(\Delta z_m))' (z_i \in \mathbf{Z}^2)$ と置き, 目的関数を

$$L_m^*(\theta; \Delta \mathbf{X}_m) = \int_{[-\pi / \Delta, \pi / \Delta]^2} \left(\log f_{\Delta X}(\lambda; \theta) + \frac{I_m(\lambda)}{f_{\Delta X}(\lambda; \theta)} \right) d\lambda$$

によって定義する. $I_m(\lambda)$ はピリオドグラム $I_m(\lambda) = |\sum_{j=1}^m X(\Delta z_j) e^{i \Delta z_j' \lambda}|^2$ である. $L_m^*(\theta; \Delta \mathbf{X}_m)$ を最小にする θ をパラメータとして採用する.

ただし実際の計算では (4.2) の無限級数は適当な有限級数で近似する. また $X(\Delta z)$ は観測できないので,

$$(4.3) \quad \tilde{X}(\Delta z) = \frac{1}{n_z} \sum_{s_i \in D_{\Delta z}} Y(s_i)$$

によって代替する. ここで $D_{\Delta z}$ は Δz を中心とする一辺の長さ Δ の正方形, n_z はそれに含まれる s_i の個数とする.

もう一つは Matsuda and Yajima (2009) によって提案された方法である. ここでは以下のようなサンプリング・スキームを考える. 観測地点は

$$s_i = (A_1 u_{i,1}, \dots, A_d u_{i,d})', \quad i = 1, \dots, n$$

とする. ここで

$$u_i = (u_{i,1}, \dots, u_{i,d})', \quad i = 1, \dots, n,$$

は $[0, 1]^d$ をサポートに持つ確率密度関数 $g(x)$ から生成される i.i.d. サンプルとする. 各辺を $A_j = A_j(k) (j=1, \dots, d)$, サンプル数を $n = n_k$ とおき $k \rightarrow \infty$ のとき発散すると仮定する.

このとき観測値のフーリエ変換およびピリオドグラムを

$$J_k(\lambda) = (2\pi)^{-\frac{d}{2}} G^{-1/2} |S_k|^{\frac{1}{2}} n_k^{-1} \sum_{p=1}^{n_k} Y_{s_p} \exp(-i s_p' \lambda)$$

$$I_k(\lambda) = |J_k(\lambda)|^2$$

によって定義する. ここで $S_k = [0, A_1] \times \dots \times [0, A_d]$, $|S_k|$ はその体積 $|S_k| = A_1 \times \dots \times A_d$ また $G = \int_{[0,1]^d} g(x)^2 dx$ とする. $J_k(\lambda)$ は $(2\pi)^{-\frac{d}{2}} |S_k|^{-1/2} \int_{S_k} Y_s \exp(-i s' \lambda) ds$ を近似する統計量である. ただし G は未知なのでカーネル関数などを用いてノンパラメトリックに推定する.

目的関数は Whittle 推定量と同じアイデアに基づき

$$(4.4) \quad L_k(\theta) = \int_D \left[\log\{f(\lambda:\theta) + c_k(\theta)\} + \frac{I_k(\lambda)}{f(\lambda:\theta) + c_k(\theta)} \right] d\lambda$$

によって定義する. ここで D は $\lambda \in D \rightarrow -\lambda \in D$ をみたす対称なコンパクト集合とする. $c_k(\theta)$ はピリオドグラムバイアスを考慮して必要となる θ の関数である. $L_k(\theta)$ を最小にする推定量は $k \rightarrow \infty$ のとき一貫性・漸近正規性をみtas.

ここで3つの方法を比較したシミュレーション結果を紹介する (Matsuda and Yajima, 2009). 自己共分散関数は

$$C(h_1, h_2) = \sigma_0 I(h_1=0, h_2=0) + \frac{\sigma_1}{2^{\nu-1} \Gamma(\nu)} \left(\frac{2\nu^{1/2} r}{\rho} \right)^\nu \mathcal{K}_\nu \left(\frac{2\nu^{1/2} r}{\rho} \right) \\ r^2 = [\beta\{h_1 \cos(\theta) - h_2 \sin(\theta)\}]^2 + [\beta^{-1}\{h_1 \sin(\theta) + h_2 \cos(\theta)\}]^2.$$

によって定義する. 右辺第1項の $\sigma_0 I(h_1=0, h_2=0)$ はナゲット効果 (nugget effect) と呼ばれ, $h_i (i=1,2) \rightarrow 0$ のとき, 自己共分散関数がジャンプする場合を考慮した量である. 第2項はパラメタライゼーションは異なるが, 2章で説明した Matérn 族に属する自己共分散関数である. 一方 r は単純な距離とはせず β, θ は非等方性を表現するパラメータである. まず座標 (h_1, h_2) を角度 θ だけ回転させ

$$\begin{pmatrix} l_1 \\ l_2 \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$$

を新たな座標とする. 次に座標 (l_1, l_2) において l_1 は β だけ拡大し, l_2 は逆に β だけ縮小する. その結果距離の等高線は楕円

$$r^2 = \beta^2 l_1^2 + \beta^{-2} l_2^2.$$

になる. $\beta=1, \theta=0$ のとき $\gamma(h_1, h_2)$ は等方型モデルになる.

このときスペクトル密度関数は

$$f(\lambda_1, \lambda_2) = \phi(\alpha^2 + \omega^2)^{-(\nu+1)}$$

となる. ここで

$$\omega^2 = [\beta^{-1}\{\lambda_1 \cos(\theta) - \lambda_2 \sin(\theta)\}]^2 + [\beta\{\lambda_1 \sin(\theta) + \lambda_2 \cos(\theta)\}]^2 \\ \alpha = 2\sqrt{\nu}/\rho, \phi = \nu\alpha^{2\nu} \sigma_1/\pi$$

とする. 真のパラメータは

$$\nu = 3.0, \sigma_1 = 10, \rho = 3.0, \sigma_0 = 3.0, \beta = 0.85, \theta = 0.3$$

と置く. サンプル地点は $[0, 40] \times [0, 40]$ 上の一様分布にしたがって抽出し, サンプル数は 2000, 繰り返し回数は 200 回とした. Matsuda and Yajima (2009) の目的関数 (4.4) の積分領域は $D = [-3\pi, 3\pi]^2$ とした.

以下の表がその結果である. 表1の上段が平均値, 下段の括弧内の数値が標準偏差である. 最後の2行は1回当たりの計算に要するCPUタイムの平均値とその標準偏差である. 表2は等方性 ($\beta=1, \theta=0$) を帰無仮説とする仮説検定に尤度比検定を適用した結果である. χ^2 分布に基づく名目的な有意水準を 0.05 に取ったときの, 実際的水準 (size) と検出力 (power) を示している.

表からは Stein et al. (2004) の方法が ν の推定を除いて最良であり, また尤度比検定における検出力も一番高い. ただし Matsuda and Yajima (2009) の方法は ν の推定に関しては最良で

表 1. 推定量の平均値と標準偏差.

	M and Y	Fuentes	Stein
$\nu(= 3.0)$	2.85 (0.40)	4.27 (1.08)	3.96 (2.13)
$\sigma_1(= 10)$	10.07 (1.44)	12.30 (1.50)	10.09 (1.31)
$\sigma_0(= 3)$	2.98 (0.87)	2.72 (0.44)	3.00 (0.15)
$\rho(= 3.0)$	2.85 (0.24)	2.26 (0.17)	3.02 (0.34)
$\beta(= 0.85)$	0.86 (0.06)	0.91 (0.09)	0.85 (0.03)
$\theta(0.3)$	0.27 (0.22)	0.17 (0.43)	0.29 (0.12)
time(sec)	8.7 4.2	53.1 21.7	189.2 47.2

表 2. 尤度比検定の水準と検出力.

	M and Y	Fuentes	Stein
Size(=0.05)	0.08	0.13	0.08
Power	0.74	0.29	0.99

ある. 理由は自己共分散関数よりスペクトル密度関数の方が ν の値が異なるとその形状の相違がより顕著になることにあると思われる. Fuentes (2007) の方法は推定・検定とも Stein et al. (2004) あるいは Matsuda and Yajima (2009) ほど良くはない. (4.2) の無限級数を有限の級数により, さらに (4.1) の積分を (4.3) の和によって各々近似することに起因すると考えられる. Matsuda and Yajima (2009) の優れた点は計算速度の速さである. ちなみに平均計算時間 (8.7 秒) は Stein et al. (2004) の方法 (189.2 秒) の 5% しか要しない.

5. 潜在過程モデル

時空間データはしばしば観測誤差を伴い, 直接には観測できない潜在的な確率場の推測が目的になる場合がある. 本章では潜在的な確率場の推測を高速に行うために提案された 2 つの方法を紹介する. まず観測される確率場 $Z(\mathbf{s})$ を

$$(5.1) \quad Z(\mathbf{s}) = Y(\mathbf{s}) + U(\mathbf{s})$$

によって定義する. ここでは潜在する確率場 $\{Y(\mathbf{s})\}$ が 3 章で述べた回帰モデル

$$(5.2) \quad Y(\mathbf{s}) = m(\mathbf{s})' \beta + \epsilon(\mathbf{s})$$

にしたがう. ただし $\epsilon(\mathbf{s})$ が非正常確率場になることも考慮してその自己共分散関数を

$$\text{Cov}(\epsilon(\mathbf{s}), \epsilon(\mathbf{s}')) = C(\mathbf{s}, \mathbf{s}')$$

と置く. $\{U(\mathbf{s})\}$ は観測誤差などを意味する期待値 0 分散 $\tau^2 v(\mathbf{s})$ の互いに無相関な確率変数列とし, $v(\mathbf{s})$ は既知の関数とする.

観測値ベクトルを $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))'$ と置く. C, V を各々 $C = (C(\mathbf{s}_i, \mathbf{s}_j)), V = \text{diag}(v(\mathbf{s}_1), \dots, v(\mathbf{s}_n))$ によって定義される $n \times n$ 行列とすれば, \mathbf{Z} の共分散行列は

$$(5.3) \quad \Sigma = C + \tau^2 V$$

になる.

データ解析を高速化するために, 多変量解析における因子分析と同様の発想に基づき共分散行列 C を次元を縮約した行列により表現する方法がある. ひとつは Banerjee et al. (2008) によって提案された正規予測過程 (Gaussian Predictive Process) であり, 他方は Cressie and Johannesson (2008) によって提案された固定階数クリギング (Fixed Rank Kriging) である.

正規予測過程はデータ観測において主要と思われる地点を $m (< n)$ 個まず選び, $D^* = (\mathbf{s}_1^*, \dots, \mathbf{s}_m^*)$ とする. 予測過程 $\{\tilde{\epsilon}(\mathbf{s})\}$ は $\epsilon(\mathbf{s}_i^*) (i = 1, \dots, m)$ が与えられたときの BLUP

$$\tilde{\epsilon}(\mathbf{s}) = c(\mathbf{s})'(C^*)^{-1}\epsilon^*$$

によって定義する. ここで C^* は $m \times m$ 行列, $c(\mathbf{s}), \epsilon^*$ は各々 m 次元ベクトルで

$$C^* = (C(\mathbf{s}_i^*, \mathbf{s}_j^*))$$

$$c(\mathbf{s}) = (c_1, \dots, c_m)', c_i = \text{Cov}(\epsilon(\mathbf{s}), \epsilon(\mathbf{s}_i^*))$$

$$\epsilon^* = (\epsilon(\mathbf{s}_1^*), \dots, \epsilon(\mathbf{s}_m^*))'$$

とする. (5.2) において $\epsilon(\mathbf{s})$ を $\tilde{\epsilon}(\mathbf{s})$ で置き換えて

$$(5.4) \quad Z(\mathbf{s}) = m(\mathbf{s})'\beta + \tilde{\epsilon}(\mathbf{s}) + U(\mathbf{s})$$

を正規予測過程と呼ぶ. (5.1) と (5.4) は同じ $Z(\mathbf{s})$ により表現しているが, 異なった確率場である.

一方固定階数クリギングでは (5.2) において $\epsilon(\mathbf{s}) = \mathbf{B}'(\mathbf{s})\eta$ と置く. ここで η は $m (< n)$ 次元確率ベクトルでその共分散行列を \mathbf{G} とする. $\mathbf{B}(\mathbf{s}) = (B_1(\mathbf{s}), \dots, B_m(\mathbf{s}))'$ は m 個の基底関数の集合 (必ずしも直交しない) である. このとき (5.3) は

$$(5.5) \quad \Sigma = \mathbf{B}\mathbf{G}\mathbf{B}' + \tau^2 V$$

となる. ここで \mathbf{B} は $n \times m$ 行列でその (i, j) 成分は $B_j(\mathbf{s}_i)$ とする. \mathbf{B} および \mathbf{G} の選択により, 多様な非定常確率場の共分散行列を表現できる. また (5.5) の Σ の逆行列は

$$\Sigma^{-1} = (\tau^2 V)^{-1} - (\tau^2 V)^{-1} \mathbf{B}(\mathbf{G}^{-1} + \mathbf{B}'(\tau^2 V)^{-1} \mathbf{B})^{-1} \mathbf{B}'(\tau^2 V)^{-1}$$

である. V は対角行列なので, 実質的に Σ^{-1} の計算は $m \times m$ 行列 $\mathbf{G}^{-1} + \mathbf{B}'(\tau^2 V)^{-1} \mathbf{B}$ の逆行列の計算に帰着する.

6. 今後の課題

最後に本章では今後解決すべき課題について列挙する.

(i) Covariance Tapering については, まず実際のデータ解析では真の自己共分散関数は未知であり, 推定量を代入した場合の影響を明らかにしなければならない.

次に Taper のレンジ θ をサンプル数に依存させてどの様にするべきかについて実用上の指針が必要である. 次に Matérn 族に代表される等方型以外の非等方型あるいは分離型自己共分散関数に対してその理論的性質を導く必要がある. たとえば分離型モデルにおいては, 自己共分散関数を 2 つに分割した場合, $C(\mathbf{h})$ から生成される観測値間の共分散行列は $C_i(\tilde{\mathbf{h}}_i) (i = 1, 2)$ から生成される共分散行列のクロネッカー積によって表現できる. したがってその逆行列も

各々の逆行列のクロネッカー積になり、大規模データの処理に都合がよいモデルである。いま $C_1(\tilde{h}_1)$ を空間自己共分散関数、 $C_2(\tilde{h}_2)$ を時間自己相関関数とすれば、空間自己共分散関数を tapering すれば、さらに高速計算が可能になる。この場合に定理 4, 5 が一般化できるかまだ未解決である。

最後に対数正規確率場など正規定常確率場の非線形変換により得られる定常確率場への拡張が考えられる。いま $\{Y(s)\}$ は期待値 0, 分散 1, 自己共分散関数は Matérn 族に属する正規確率場とする。実数値関数 $T(x): \mathbf{R} \rightarrow \mathbf{R}$ を用いて、新たな確率場 $\{Z(s)\}$ を

$$Z(s) = T(Y(s)),$$

によって定義する。たとえば $T(x) = \exp(x)$ とおけば、 $\{Z(s)\}$ は対数正規確率場になる。 $\phi(x)$ を標準正規分布の密度関数 $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ としたとき、 $T(x)$ が

$$\int_{-\infty}^{\infty} T(x)^2 \phi(x) dx < \infty,$$

をみたすならば、 $Z(s)$ はエルミート多項式の無限和

$$Z(s) = \sum_{j=1}^{\infty} a_j H_j(Y(s)),$$

によって表現できる。ここで $H_j(x) (j=1, 2, \dots)$ は j 次のエルミート多項式であり、右辺は平均二乗収束による無限和である。 $Z(s)$ の自己共分散関数とスペクトル密度関数は

$$C_Z(\|\mathbf{h}\|) = \sum_{j=1}^{\infty} a_j^2 j! C_Y(\|\mathbf{h}\|)^j$$

$$f_z(\|\boldsymbol{\lambda}\|) = \sum_{j=1}^{\infty} a_j^2 j! f_Y^{(j)}(\|\boldsymbol{\lambda}\|),$$

となる。ここで $f_Y^{(j)}(\|\boldsymbol{\lambda}\|)$ は $f_Y(\|\boldsymbol{\lambda}\|)$ の j 回結合関数である。

このような確率場についても定理 5 が成立すると予想される (Hirano, 2011)。

(ii) 本稿で紹介した 3 つの尤度関数に対する近似法についても各々今後考察すべき課題がある。まず Stein et al. (2004) に関しては、前述のごとくシミュレーションにおいては良好なパフォーマンスを示すが、その理論的正当性は導出されていない。また観測値の順序づけ、グループへの分割方法及びその個数について実際的なガイドラインが必要である。

Fuentes (2007) については正方形の一辺の長さ Δ の決定方法および (4.2) の無限級数を有限級数で切断する方法についてより詳細に議論する必要がある。

最後に Matsuda and Yajima (2009) においては、推定量の精度は積分領域 D の選択に依存する。サンプルを無駄にしないためにはサンプル数が増大するとともに領域 D も拡張していくことが自然に思われるが、どのような速度で増大させれば一貫性および漸近正規性が依然として成立するかまだ未解決である。

(iii) 正規予測過程については地点 $D^* = (s_1^*, \dots, s_m^*)$ およびその個数 m の選択方法の開発が必要となる。一方固定階数クリギングについては基底関数 $\mathbf{B}(s) = (B_1(s), \dots, B_m(s))'$ の選択方法とその個数 m の選択方法の開発が必要となる。同時に両方法の長所、短所、互いの優劣なども今後解明されなければならない。

謝 辞

初稿を改訂するにあたり、査読者から大変有益なご助言をいただいた。ここに謝意を表したい。

参 考 文 献

- Banerjee, S., Gelfand, A. E., Finley, A. O. and Sang, H. (2008). Gaussian predictive process models for large spatial data sets, *Journal of the Royal Statistical Society*, **B70**, 825–848.
- Cressie, N. (1993). *Statistics for Spatial Data*, rev. ed., Wiley, New York.
- Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large data sets, *Journal of the Royal Statistical Society*, **B70**, 209–226.
- Davis, T. A. (2006). *Direct Methods for Linear Sparse Systems*, SIAM, Philadelphia.
- Du, J., Zhang, H. and Mandrekar, V. S. (2009). Fixed domain asymptotic properties of tapered maximum likelihood estimators, *The Annals of Statistics*, **37**, 3330–3361.
- Fuentes, M. (2007). Approximate likelihood for large irregularly spaced spatial data, *Journal of the American Statistical Association*, **102**, 321–331.
- Furrer, R., Genton, M. and Nychka, D. (2006). Covariance tapering for interpolation of large spatial data sets, *Journal of Computational and Graphical Statistics*, **15**, 502–523.
- Guttorp, P. and Gneiting, T. (2006). Studies in the history of probability and statistics XLIX, On the Matérn correlation family, *Biometrika*, **93**, 989–995.
- Guyon, X. (1995). *Random Fields on a Network, Modeling, Statistics and Applications*, Springer, New York.
- Hirano, T. (2011). Covariance tapering for prediction of large spatial data sets in transformed random fields (in preparation).
- Kaufman, C. G., Schervish, M. J. and Nychka, D. W. (2008). Covariance estimation for likelihood-based estimation in large spatial data sets, *Journal of the American Statistical Association*, **103**, 1545–1555.
- Matérn, B. (1947). *Metoder att Uppskatta Noggranheten vid Linje- och Provytetaxering*, Stockholm: Medd Statens Skogsforskningsinstitut, **36**, no. 1, (In Swedish with substantial English summary).
- Matérn, B. (1960). *Spatial Variation-stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations*, Stockholm: Medd. Statens Skogsforskningsinstitut, **49**, no. 5.
- Matérn, B. (1986). *Spatial Variation*, 2nd ed., Lecture Notes in Statistics, Springer, Berlin.
- Matsuda, Y. and Yajima, Y. (2009). Fourier analysis of irregularly spaced data on R^d , *Journal of the Royal Statistical Society*, **B71**, 191–217.
- Putter, H. and Young, G. A. (2001). On the effect of covariance function estimation on the accuracy of kriging predictors, *Bernoulli*, **7**, 421–438.
- Stein, M. L. (1990). Uniform asymptotic optimality of linear prediction of a random field using an incorrect second-order structure, *The Annals of Statistics*, **18**, 850–872.
- Stein, M. L. (1993). A simple condition for asymptotic optimality of linear prediction of random fields, *Statistics and Probability Letters*, **17**, 399–404.
- Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, New York.
- Stein, M. L., Chi, Z. and Welty, L. J. (2004). Approximate likelihood for large spatial datasets, *Journal of the Royal Statistical Society*, **B66**, 275–296.
- Su, Y., Li, B. and Genton, M. (2011). Geostatistics for large data sets, *Space-Time Processes and Challenges Related to Environmental Problems: Proceedings of the Spring School “Advances and Challenges in Space-time Modelling of Natural Events”* (eds. E. Porcu, J. M. Montero and M. Schlather), Springer, Berlin (to appear).
- Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes, *Journal of the Royal Statistical Society*, **B50**, 297–312.

- Wang, D. and Loh, W-L. (2011). On fixed-domain asymptotics and covariance tapering in Gaussian random fields models, *Electronic Journal of Statistics*, **5**, 238–269.
- Yaglom, Y. (1987). *Correlation Theory of Stationary and Related Random Functions*, Vol. I, Springer, New York.
- 矢島美寛 (2011). 時系列解析から時空間統計解析への展望, 日本統計学会誌, **41**, シリーズ J, 219–244.

Statistical Analysis of Large Spatio-temporal Data Sets

Yoshihiro Yajima and Toshihiro Hirano

Graduate School of Economics, The University of Tokyo

We review various time-saving spatio-temporal statistical methodologies and discuss problems to be solved in future. First we consider covariance tapering for the best linear unbiased estimator (BLUP), which is called kriging in geostatistics. Secondly we consider likelihood approximation in both spatio-temporal and frequency domains. Thirdly we describe latent process models which reduce the number of the parameters substantially so that it is able to analyze large spatio-temporal data sets within feasible computational time. Finally we discuss open problems to be solved in future