

時系列解析から時空間統計解析への展望¹

矢島 美寛*

From Time Series Analysis to Spatio-Temporal Statistical Analysis

Yoshihiro Yajima*

時系列解析および時系列解析を一般化した時空間統計解析の過去から現在までの進展と将来への展望について議論する。

We review developments of time series analysis and spatio-temporal statistical analysis and envisage a future of them.

キーワード: 時系列解析, 時空間統計解析

1. 序

本稿では, データ間の時間的な相互作用あるいは因果関係を的確に表現するモデルの構築を目的とする時系列解析からさらには空間的なダイナミズムも考慮したモデルの構築を目的とする時空間統計解析について現在までの進展と将来への展望について概説する. ただし関係するトピックは広範な分野に渡り, 内容は網羅的あるいは中立的ではなく, 選択的である. あくまで筆者の能力で叙述できる範囲および興味の範囲に限られていることをあらかじめ断っておく. 時空間統計解析を一望できる文献として Gelfand *et al.* (2010) がある. また Finkenstät *et al.* (2007) は入門的な内容から最近の発展まで平易に解説している.

2 節以降の構成は以下の通りである. 2 節では時系列解析の歴史について簡単に振り返る. 3 節では時系列解析に用いられる確率過程のうち中心的な役割を果たしてきた定常過程の定義を与える. 次に系列相関の強さによって定常過程を二分して短期記憶定常過程と長期記憶定常過程の定義を与える. この定義の下で両過程に対する代表的なモデルを紹介する. 4 節では定常過程に対するパラメトリックモデルのパラメータ推定法およびその漸近的性質について説明する. 5 節では最初に時空間統計解析について, 特に時系列解析との相違に留意しつつ説明する. 時空間統計解析において確率変数の集合は確率場と呼ばれ

¹ 本稿は日本統計学会春季集会 (2011 年 3 月 6 日) において講演した内容を加筆修正したものである.

* 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 (E-mail: yajima@e.u-tokyo.ac.jp).

る。次に確率場の一つで定常過程を一般化した定常確率場の定義を与える。6節では定常確率場に対するモデルについて説明する。7節では4節と対比しながら定常確率場に対するパラメトリックモデルの推定について説明する。最後に8節では時空間統計解析における今後の課題について列挙する。

なお叙述を簡潔にするため、定義および定理などにおいて数学的に厳密な条件を省略している箇所もある。それらについては原典を参照いただきたい。

2. 時系列解析の歴史

20世紀中葉以降の確率論および確率過程論の発展と相まって、時系列解析は時系列データのある確率過程からの実現値と見なし、時間的変動や相互作用を的確に表現するモデルの構築を目的としている。

実際のモデル構築法を確立し、それを広く知らしめる上で、Box and Jenkins (1970)の果たした貢献は大きい。彼等は自己回帰移動平均モデル (Autoregressive Moving Average model, 以下 ARMA モデルと略) と呼ばれる時系列モデルに対する同定法、推定法、適合度検定、そして予測や制御などへの応用に至る手順を体系化し、実際の時系列データを解析する人々にきわめて有効な解析方法を提供した。今日この解析法は、「Box-Jenkins 法」と呼ばれている。なお同時代に発刊され ARMA モデルを初めとする時系列解析の理論を解説した本としては、Hannan (1970), Anderson (1971), Fuller (1976), Priestley (1981) などがある。

次節で具体的表現を示すが ARMA モデルが持つ2大特徴は「線形性」と「短期記憶性」である。そのため数学的な表現は簡潔となり、理論的性質を明解に導くことができる。またパラメータの推定においては推定量の効率的な計算方法も開発されており (Brockwell and Davis (1991), Chapter 8), さらに推定量から導出された残差項の分析を通じてデータに対する適合度判定も容易である。しかし一方で急激な変化が生じたりあるいは上下動が非対称などの「非線形性」を示すデータあるいは時点が大きく隔たった観測値間の相関係数が無視できないような「長期記憶性」を示すデータに対するモデルとしては適当でない。

1980年代以降現在に至る時系列解析研究の重心は非線形性あるいは長期記憶性さらには非定常性を示すデータに適合するモデルの開発、それらのモデルが持つ理論的性質の導出、応用可能性へと移行している。前者に関しては、パラメトリックモデルとしては非線形自己回帰モデル、自己回帰条件付き分散不均一モデル (Autoregressive Conditional Heteroskedasticity model), 確率的ボラティリティ・モデル (Stochastic Volatility model) があり、さらにセミパラメトリック、ノンパラメトリック・モデルについても盛んに議論されている (Tong (1990), Taylor (1994), Guouriéroux (1997), Fan and Yao (2003), 刈屋他 (2003)). 後者に関しては長期記憶モデル、単位根モデル、共和分分析などのトピック

がある (Banerjee *et al.* (1993), Beran (1994), Doukhan *et al.* (2002), Johansen (1996), 刈屋他 (2003), Robinson (2003)).

その中で次節以降では長期記憶モデルについて解説する。

3. 定常過程とそのモデル

時系列データを解析する確率過程として中心的な役割を果たしてきたのは定常過程である。時間軸の平行移動に対して確率構造が不変な確率過程を定常過程と言う。多くの場合実際の時系列データは長時間観測すれば非定常性を示す。しかし非定常な時系列データに対するモデルであっても、内在的には何らかの定常性を仮定しない限り、意味のある安定的構造を検出することは不可能である。また応用上も有効なモデルとはなり得ない。例えば本稿では省略するが上述の単位根モデルは相隣接する 2 時点で観測されたデータの差分を取った系列が定常過程に従う。共和分モデルでは多変量時系列の個々の成分時系列は非定常過程であるが、その線形結合は定常過程に従う。

要求する不変性の強弱により強定常過程 (strictly stationary process) と弱定常過程 (weakly stationary process) がある。ここでは以下の議論で必要となる弱定常過程の定義のみを与える (Brockwell and Davis (1991))。いま時点を t で表し、 t の動く範囲を実数の集合 $\mathbf{R} = (-\infty, \infty)$ あるいは整数の集合 $\mathbf{Z} = \{0, \pm 1, \pm 2, \dots\}$ とする。統一的には \mathbf{K} と書く。

定義 3.1 弱定常過程

確率過程 $\{X_t | t \in \mathbf{K}\}$ が以下の 2 条件を満たすとき、弱定常過程に従うと言う。

- (i) $\{X_t\}$ の期待値は時点 t に依存せず一定の値 $E(X_t) = \mu$ を取る。以下簡単のため $\mu = 0$ とする。
- (ii) $Cov(X_t, X_s)$ は時間差 $t - s$ のみに依存する。

このとき自己共分散関数 $\{\gamma(h)\}$ を $\gamma(h) = Cov(X_{t+h}, X_t)$ によって定義する。以下では断りのない限り弱定常過程を簡単のため定常過程と呼ぶ。定常過程とその自己共分散関数はスペクトル表現と言うフーリエ変換によって表現できる。まず自己共分散関数は

$$\gamma(h) = \int_T \exp(ih\lambda) dF(\lambda) \quad (i^2 = -1) \quad (3.1)$$

によって表現できる。ここで $\mathbf{K} = \mathbf{R}$ のときは $T = (-\infty, \infty)$, $\mathbf{K} = \mathbf{Z}$ のときは、 $T = [-\pi, \pi]$ とする。また $F(\lambda)$ は単調非減少な右連続関数である。 $F(\lambda)$ が絶対連続な場合には、(3.1) は非負の値を取る可積分関数 $f(\lambda)$ により

$$\gamma(h) = \int_T \exp(ih\lambda) f(\lambda) d\lambda$$

と表現される. $F(\lambda)$, $f(\lambda)$ を各々スペクトル分布関数, スペクトル密度関数と言う. 一方定常過程自身は, 確率積分

$$X_t = \int_T \exp(it\lambda) dZ(\lambda) \quad (3.2)$$

によって表現される. ここで $\{Z(\lambda)\}$ は複素数値直交増分過程で任意の $\lambda_1 \leq \lambda_2$ ($\lambda_1, \lambda_2 \in T$) に対して,

$$E|Z(\lambda_2) - Z(\lambda_1)|^2 = \int_{\lambda_1}^{\lambda_2} dF(\lambda)$$

を満たす. 直観的に言うと, 定常過程は位相・振幅を表す $dZ(\lambda)$ が確率変数となるさまざまな周波数 λ の波が合成されたものである. 周波数 λ の波の強さ $E|dZ(\lambda)|^2$ は $dF(\lambda)$ あるいは $f(\lambda)d\lambda$ に等しい. 以下では $\mathbf{K} = \mathbf{Z}$ とする.

次に定常過程を自己共分散関数の強さにより 2 つのクラスに分類する. 自己共分散関数が

$$\sum_{h=0}^{\infty} |\gamma(h)| < \infty \quad (3.3)$$

を満たすとき, $\{X_t\}$ を短期記憶 (short-memory) あるいは短期依存 (short range dependent) あるいは弱従属 (weakly dependent) 定常過程と言う. 逆に

$$\sum_{h=0}^{\infty} |\gamma(h)| = \infty \quad (3.4)$$

となるとき, 長期記憶 (long-memory) あるいは長期依存 (long range dependent) あるいは強従属 (strongly dependent) 定常過程と言う.

次に短期および長期記憶定常過程に対する代表的なモデルを挙げる. まず短期記憶定常過程に対するモデルとしては, 前述の ARMA モデルがある. $\{X_t\}$ が

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = U_t - \theta_1 U_{t-1} - \dots - \theta_q U_{t-q} \quad (3.5)$$

を満たすとき, $\{X_t\}$ は次数 (p, q) の ARMA モデルに従うと言う. ここで $\{U_t\}$ は白色雑音 (white noise) と呼ばれる分散一定 $\text{Var}(U_t) = \sigma^2$, 互いに無相関な確率変数列 $\text{Cov}(U_t, U_s) = 0$ ($t \neq s$) である. $\{U_t\}$ のスペクトル密度関数は周波数に依存せず一定で

$$f_U(\lambda) = \frac{\sigma^2}{2\pi} \quad (3.6)$$

となる. いま後退作用素 B を $BX_t = X_{t-1}$ によって定義すれば, (3.5) はコンパクトに

$$\phi(B)X_t = \theta(B)U_t$$

によって表現できる. ここで

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$$

とする。 $\phi(z) = 0$ の根が単位円周上 ($|z| = 1$) に存在しなければ ARMA モデルは定常過程になる。さらにすべての根が単位円の外側に存在すれば X_t は過去と現在の白色雑音 $U_s (s \leq t)$ の線形結合によって表現できる。これを因果性 (causality) を満たすための条件と言う。自己共分散関数は時間差 h が増大するとともに指数的に 0 へ減衰し、不等式

$$|\gamma(h)| \leq K a^{|h|} \quad (0 < a < 1) \quad \forall h \tag{3.7}$$

を満たす。ここで K は正定数とする。従って (3.3) が成立し、短期記憶モデルである。一方スペクトル密度関数は、三角関数の比となる有理型関数

$$f(\lambda) = \frac{\sigma^2 |\theta(e^{i\lambda})|^2}{2\pi |\phi(e^{i\lambda})|^2} \tag{3.8}$$

となり、滑らかな無限回微分可能な関数である。フーリエ解析における基本的な定理より関数が滑らかであればあるほどそのフーリエ変換は迅速に 0 へ収束する。(3.7) と (3.8) はその関係を表している。

次に長記憶定常過程に対するモデルを 2 つ紹介する。一つは ARMA モデルを一般化した ARFIMA モデルである (Granger and Joyeux (1980), Hosking (1981))。"FI" は fractionally integrated の略である。これの意味するところを明らかにするために、まず任意の実数 d に対して実数差分作用素 (fractional difference operator) ∇^d を、一般化された二項係数を用いて

$$\begin{aligned} \nabla^d &= (1 - B)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-B)^j \tag{3.9} \\ \binom{d}{j} &= \frac{d(d-1)\dots(d-j+1)}{j!} = \frac{\Gamma(d+1)}{\Gamma(j+1)\Gamma(d-j+1)} \end{aligned}$$

によって定義する。ここで $\Gamma(x)$ はガンマ関数である。 d が非負整数であれば $\binom{d}{j} = 0 (j > d)$ となり、(3.9) は有限級数であるから通常差分作用素を意味する。 d が実数の場合もヒルベルト空間の作用素論を応用して数学的に正当化できる (Yajima (1985))。

∇^d を用いて、 $\{X_t\}$ が

$$\phi(B)\nabla^d X(t) = \theta(B)U_t$$

を満たすとき、 $\{X_t\}$ は次数 (p, d, q) の ARFIMA モデルに従うと言う。 $d < 1/2$ のとき $\{X_t\}$ は定常過程になり、そのスペクトル密度関数は

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{i\lambda})|^2}{|\phi(e^{i\lambda})(1 - e^{i\lambda})^d|^2} \tag{3.10}$$

となる。ARMA モデルのスペクトル密度関数 (3.8) と比べると、(3.10) の分母に ∇^d の効果 $(1 - e^{i\lambda})^d$ が現れている。従って $0 < d < 1/2$ のとき、 $\lambda \rightarrow 0$ とすれば $f(\lambda) \rightarrow \infty$ となる。すなわち低周波の波言い換えると周期の長い波の優勢な長期記憶性を持つモデルである。

2番目のモデルはフラクショナル・ガウシアン・ノイズ (Fractional Gaussian Noise, 以下 FGN モデルと略) と呼ばれる (Manderbrot and van Ness (1968)). このモデルを導入するための準備としてフラクショナル・ブラウン運動 (Fractional Brownian Motion, 以下 FBM モデルと略) を定義する. 正規確率過程 $\{B_H(t)|0 \leq t < \infty\}$ が以下の2条件

$$\begin{aligned} E(B_H(t)) &= 0 \quad \forall t \\ E(|B_H(t) - B_H(s)|^2) &= \sigma^2 |t - s|^{2H} \quad (0 < H < 1) \end{aligned}$$

を満たすとき, FBM モデルに従うと言う. $H = 1/2$ のときは通常のブラウン運動になる.

FBM モデルを用いて, 確率過程 $\{X_t|t = 1, 2, \dots\}$ を

$$X_t = B_H(t) - B_H(t-1)$$

によって定義する. $\{X_t\}$ を FGN モデルと呼ぶ. $H = 1/2$ (ブラウン運動) の場合, FGN モデルは互いに独立で同一な分布 $N(0, \sigma^2)$ に従う確率変数列であるが, $H \neq 1/2$ のときは系列相関を持つ定常過程になる. 特に $1/2 < H$ の場合, 陽には表現できないがスペクトル密度関数は ARFIMA モデルのそれと同じく $\lambda \rightarrow 0$ のとき無限大に発散する (Sinai (1976)).

ARFIMA モデルと FGN モデルのスペクトル密度関数および自己共分散関数は統一的に

$$\begin{aligned} f(\lambda) &\sim |\lambda|^{\alpha-1} \quad (\lambda \rightarrow 0) \\ \gamma(h) &\sim h^{-\alpha} \quad (h \rightarrow \infty) \end{aligned}$$

によって表現できる. ここで

$$\alpha = 1 - 2d = 2 - 2H \tag{3.11}$$

である. 従って $0 < \alpha < 1$ ($0 < d < 1/2$, $1/2 < H < 1$) の場合, スペクトル密度関数は

$$f(\lambda) \rightarrow \infty \quad (\lambda \rightarrow 0)$$

となり, 一方自己共分散関数については (3.4) が成立する.

4. パラメトリックモデルのパラメータ推定

定常過程の統計的推測理論に関しては, スペクトル密度関数あるいは回帰モデルにおける推定・検定問題など多くのトピックがある. 紙幅の関係で本稿ではパラメトリック・モデルのパラメータ推定について取り上げる.

$\{X_t\}$ が正規定常過程に従う場合には最尤推定量をパラメータ推定に应用することができる. いまパラメータ・ベクトルを θ とする. 例えば ARMA モデルにおいては係数 $\phi_i (i = 1, \dots, p), \theta_j (j = 1, \dots, q)$ および $\sigma^2 = \text{Var}(U_t)$ が, ARFIMA モデル, FGN モデ

ルにおいては d , H などがその成分である. $f(\lambda; \theta)$, $\gamma(h; \theta)$ をパラメータ θ のもとの, 各々スペクトル密度関数, 自己共分散関数とする. いま観測値を $\mathbf{X}_N = (X_1, \dots, X_N)'$ とする. ここで $'$ は転置ベクトルを意味する. このとき対数尤度関数は定数項を除くと

$$\log L_N(\theta; \mathbf{X}_N) = -\frac{1}{2} \log \det(\Gamma_N) - \frac{1}{2} (\mathbf{X}_N - \boldsymbol{\mu})' \Gamma_N^{-1} (\mathbf{X}_N - \boldsymbol{\mu})$$

になる. ここで $\boldsymbol{\mu} = (\mu, \mu, \dots, \mu)'$, また Γ_N は \mathbf{X}_N の共分散行列でその (i, j) 成分は $Cov(X_i, X_j) = \gamma(i - j; \theta)$ とする.

しかし時系列データの場合 Γ_N^{-1} あるいは $\det(\Gamma_N)$ の計算は複雑で長時間を要する. そこで尤度関数を近似しかつ高速計算可能な目的関数を作り, それを最適化する θ を推定量として採用する方法が提案されている. 最もポピュラーな推定量の一つが Whittle 推定量である. 簡単のため $\mu = 0$ を仮定し, 以下の近似式を用いる.

$$\begin{aligned} & -2(2\pi/N) \log L_N(\theta; \mathbf{X}_N) \\ &= (2\pi/N) [\log \det \Gamma_N + \mathbf{X}'_N \Gamma_N^{-1} \mathbf{X}_N] \\ &\sim (2\pi/N) \left[\sum_{j=1}^N \log \tau_{j, N} + \sum_{m, n=1}^N \int_{[-\pi, \pi]} \frac{\sum_{m, n=1}^N X_m X_n e^{i(m-n)\lambda}}{(2\pi)^2 f(\lambda; \theta)} d\lambda \right] \end{aligned} \quad (4.1)$$

$$\begin{aligned} &\sim \int_{[-\pi, \pi]} \left(\log f(\lambda; \theta) + \frac{I_N(\lambda)}{f(\lambda; \theta)} \right) d\lambda \quad (4.2) \\ &= L_N^*(\theta; \mathbf{X}_N) \text{ (say)}. \end{aligned}$$

ここで $I_N(\lambda)$ はピリオドグラム $I_N(\lambda) = \frac{|\sum_{t=1}^N X_t \exp(-it\lambda)|^2}{2\pi N}$, $\tau_{i, N}$ は Γ_N の固有値である. $L_N^*(\theta; \mathbf{X}_N)$ を最小にする θ を Whittle 推定量と言う.

近似式 (4.1) は Γ_N^{-1} に行列 $\Gamma_N^* = (\gamma^{mn})$ を代入することにより導かれる. ここで

$$\gamma^{mn} = \int_{[-\pi, \pi]} \frac{e^{i(m-n)\lambda}}{(2\pi)^2 f(\lambda; \theta)} d\lambda$$

とする. 厳密に言えば $\Gamma_N^* - \Gamma_N^{-1}$ は非負定値行列であるが (Shaman (1976)), (3.6) から分かるように白色雑音の場合には両行列は $\Gamma_N^{-1} = \Gamma_N^* = I_N / \sigma^2$ となり一致する. I_N は $N \times N$ 単位行列である.

一方 $f(\lambda)$ が有界でかつ 0 から有界であるとき, 任意の連続関数 $G(x)$ に対して

$$\lim_{N \rightarrow \infty} (2\pi/N) \sum_{j=1}^N G(\tau_{j, N}) = \int_{[-\pi, \pi]} G(f(\lambda)) d\lambda \quad (4.3)$$

が成立する (Grenander and Szegö (1984)). (4.3) において $G(x) = \log x$ とすれば近似式 (4.2) を得る.

Whittle 推定量の漸近的性質については, 数多くの文献がある. そのいくつかを選択的・年代順に挙げておく. まず短期記憶定常過程については

- Idea: Whittle (1961)
- Univariate Short-Memory: Walker (1964), Hannan (1973)
- Multivariate Short-Memory: Dunsmuir (1979)
- Univariate Short-Memory (Irregularly Spaced): Dunsmuir and Robinson (1981)
- Multivariate Short-Memory (Misspecified): Hosoya and Taniguchi (1982)

などが論じている.

次に長期記憶定常過程については

- Long-Memory Gaussian: Yajima (1985), Fox and Taqqu (1986)
- Long-Memory Gaussian (Misspecified): Yajima (1992)
- Long-Memory Non Gaussian: Giraitis and Surgailis (1990), Hosoya (1997)

がある.

主な結果は以下の通りである. 表現を簡潔にするため正規性を仮定するが, 非正規な場合は高次のキュムラントが極限分布に現れる以外は同様の結果が成立する.

まず $f(\lambda)$ を真のスペクトル密度関数とし,

$$\int_{[-\pi, \pi]} \left(\log f(\lambda; \theta) + \frac{f(\lambda)}{f(\lambda; \theta)} \right) d\lambda$$

を最小にする θ を θ_0 とおく. モデルが識別可能性を満たし正しく特定化されている場合, すなわち $f(\lambda) = f(\lambda; \theta), \forall \lambda$ を満たす θ が一意に存在する場合には θ_0 が真のパラメータとなる. このとき以下の定理が短期記憶定常過程あるいは長期記憶定常過程にかかわらず成立する.

定理 4.1 $N \rightarrow \infty$ とする.

- $\hat{\theta}$ は強一致性を満たし, $\hat{\theta} \rightarrow \theta_0$ *a.s.* が成立する.
- $\sqrt{N}(\hat{\theta}_N - \theta_0)$ の極限分布は $N(0, 4\pi W(\theta_0)^{-1})$ となる. ここで

$$W(\theta) = \int_{[-\pi, \pi]} \left[\frac{\partial h(\lambda; \theta)}{\partial \theta} \right] \left[\frac{\partial h(\lambda; \theta)}{\partial \theta} \right]' f(\lambda)^2 d\lambda$$

また $h(\lambda; \theta) = f^{-1}(\lambda; \theta)$ とする.

一方モデルの特定化を誤り, $f(\lambda) \neq f(\lambda; \theta), \forall \theta$ のときは次の定理が成立する. ただし当てはめるモデルのスペクトル密度関数 $f(\lambda; \theta)$ は短期記憶モデルのそれとする.

定理 4.2 $N \rightarrow \infty$ とする.

(i) 短期記憶定常過程および長期記憶定常過程に対して, $\hat{\theta} \rightarrow \theta_0$ a.s. が成立する.

(ii)(a) 短期記憶定常過程あるいは長期記憶定常過程で (3.11) において $1/2 < \alpha < 1$ のときは, $\sqrt{N}(\hat{\theta}_N - \theta_0)$ の極限分布は $N(0, 4\pi V(\theta_0)^{-1}W(\theta_0)V(\theta_0)^{-1})$ となる. ここで

$$V(\theta) = \int_{[-\pi, \pi]} \left[\frac{\partial^2 h(\lambda; \theta)}{\partial \theta \partial \theta'} \right] f(\lambda) d\lambda$$

とする.

(b) 長期記憶過程で $\alpha = 1/2$ のときは, $\sqrt{N/\log N}(\hat{\theta}_N - \theta_0)$ の極限分布は共分散行列の階数が 1 の退化した多変量正規分布になる.

(c) 長期記憶過程で $0 < \alpha < 1/2$ のときは, $(N^\alpha/\log N)(\hat{\theta}_N - \theta_0)$ の極限分布は非正規分布になる.

定理 4.1, 4.2 について 2 点注意をしておく. まずモデルを正しく特定化した場合, 定理 4.2(ii)(a) において $f(\lambda) = f(\lambda; \theta_0)$ となり, パラメタライゼーションに対するある仮定の下で $V(\theta_0) = W(\theta_0)$ が成立する (Hosoya and Taniguchi (1982)). 従って定理 4.2(ii)(a) は定理 4.1(ii) に帰着する. 次に定理 4.1(ii), 4.2(ii) の相違について考える. Whittle 推定量の極限分布は \mathbf{X}_N の 2 次形式

$$Q_N = \mathbf{X}'_N B_N \mathbf{X}_N - E(\mathbf{X}'_N B_N \mathbf{X}_N)$$

の極限分布に帰着される. ここで $B_N = (b_{mn})$ は $N \times N$ 行列でその (m, n) 成分は実数値可積分な偶関数 $\hat{b}(\lambda) (-\pi \leq \lambda \leq \pi)$ のフーリエ変換

$$b_{mn} = \int_{[-\pi, \pi]} e^{i(m-n)\lambda} \hat{b}(\lambda) d\lambda$$

によって定義される. このとき $f(\lambda)\hat{b}(\lambda)$ が $[-\pi, \pi]$ において 2 乗可積分関数であれば, $\sqrt{N}Q_N$ の極限分布は, 期待値 0, 分散が $\int_{[-\pi, \pi]} (f(\lambda)\hat{b}(\lambda))^2 d\lambda$ の定数倍に等しい正規分布になる (Fox and Taquq (1987), Giraitis and Surgailis (1990)). 一方 2 乗可積分関数でないときは, 極限分布が正規分布であっても収束の速度が $1/\sqrt{N}$ より遅くなる. あるいは極限分布が非正規分布になる場合も生じる. これを非中心極限定理 (Noncentral limit theorem) と言う (Fox and Taquq (1985)).

Whittle 推定量の場合は, ベクトル $\partial h(\lambda; \theta)/\partial \theta$ の各成分に $\theta = \theta_0$ を代入した関数が $\hat{b}(\lambda)$ に等しい. 短期記憶モデルでは $f(\lambda), \hat{b}(\lambda)$ とともに連続関数であるから, $f(\lambda)\hat{b}(\lambda)$ は常に 2 乗可積分関数になる. 一方長期記憶モデルに関しては, 定理 4.1(ii) においては $\hat{b}(0) = 0$ であるから $f(\lambda) \rightarrow \infty (\lambda \rightarrow 0)$ と相殺されることにより, また定理 4.2(ii)(a) においては $f(\lambda)$ 自身が 2 乗可積分関数でありかつ $\hat{b}(\lambda)$ は連続関数であることにより $f(\lambda)\hat{b}(\lambda)$ は 2 乗可積

分関数になる。しかし定理 4.2.(ii)(b), (c) においては $\hat{b}(0) \neq 0$ であるから 2 乗可積分関数にはならない。これらの差異が漸近分布の相違となって現れる。

当初筆者は長期記憶定常過程においては、系列相関が短期定常過程に比べ強くなり定理 4.1.(ii) は成立しないと予想していたので少なからずサプライズを覚えた。その後 Fox and Taqqu (1987) を読み、収束の速度が $1/\sqrt{N}$ となり、また極限分布が正規分布になるのは、定常過程が短期記憶過程あるいは長期記憶過程に従うかにより決定されるのではなく、関数 $f(\lambda)\hat{b}(\lambda)$ が 2 乗可積分関数か否かが本質的であることを理解した。この事実を見出した彼等の慧眼に敬服する。なおこれらの結果は次節で解説する定常確率場に対しても一般化されている。Lavancier and Philippe (2011) およびその参考文献表に所収の論文を参照されたい。

5. 時系列解析から時空間統計解析へ

時系列解析と時空間統計解析を対比的にまとめると以下のようなになる。

時系列解析

- One-parameter Stochastic Process $\{X_t | t \in \mathbf{R} \text{ あるいは } \mathbf{Z}\}$
- 自然な順序の導入が可能 (時間順序)
- 時間相関, 時間変動のモデル化

時空間統計解析

- Multi-parameter Stochastic Process $\{X_t | t \in \mathbf{R}^d \text{ あるいは } \mathbf{Z}^d\} (2 \leq d)$
- 自然な順序の導入が難しい
- 時空間相関, 時空間変動をどのようにモデル化するか?

「自然な」順序が重要である。半空間 (half space) という概念を用いて、 \mathbf{Z}^d に反射法則・対称法則・推移法則を満足する全順序を導入することは可能である (Guyon (1995))。例えば辞書式順序がある。またこの順序に基づいて t が \mathbf{Z}^d の要素である ARMA モデルを数学的正当性を持つように定義することも可能である (Guyon (1995))。しかしこのような ARMA モデルが時系列解析のように実際の時空間データの解析にどれほど関連性を持ち有効かは別問題である。

ここで以降で用いる時空間データの数学的表現を与えておく。パラメータは上述のように \mathbf{R} の d 次元直積空間あるいは $\mathbf{Z} = \{0, \pm 1, \pm 2, \dots\}$ の d 次元直積集合とする。統一的には \mathbf{K}^d と書く。観測地点 (site) は $\mathbf{s} (\in \mathbf{K}^d)$ 、観測データは $Y(\mathbf{s})$ と記す。1 変量データと多変量データの場合があるが以下では 1 変量データとする。例えば $d = 2$ のときは、 \mathbf{s} は緯

度・経度そして $Y(\mathbf{s})$ はその地点の地価などとする。 $d = 3$ であれば、 \mathbf{s} は緯度、経度、高さそして $Y(\mathbf{s})$ はその地点における気温などとする。観測時点も考慮する場合は $d = 4$ とするか、あるいは時点 t を強調するために $Y(\mathbf{s}, t)$ と表す。 $D(\subset \mathbf{K}^d)$ を \mathbf{s} の動く領域としたとき、データの全体 $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ を確率場 (random field) と言う。

確率場の一つとして、定常過程を一般化した定常確率場を定義することができる。

定義 5.2 定常確率場

$\{Y(\mathbf{s}) : \mathbf{s} \in \mathbf{K}^d\}$ が以下の 2 条件を満たすとき、定常確率場に従うと言う。

- (i) $\{Y(\mathbf{s})\}$ の期待値は \mathbf{s} に依存せず一定の値 $E(Y(\mathbf{s})) = \mu$ を取る。以下では簡単のため $\mu = 0$ とする。
- (ii) 共分散はベクトル差 $\mathbf{t} - \mathbf{s}$ のみに依存する。

$C(\mathbf{t} - \mathbf{s}) = E(Y(\mathbf{t})Y(\mathbf{s})), \mathbf{t}, \mathbf{s} \in \mathbf{K}^d$ とおき、 $\{C(\mathbf{h}), \mathbf{h} = \mathbf{t} - \mathbf{s} \in \mathbf{K}^d\}$ を弱定常過程の場合と同様に自己共分散関数と呼ぶ。 $d = 1$ の場合は時系列解析における弱定常過程である。

定常確率場および自己共分散関数は (3.1), (3.2) に対応する多変数フーリエ変換

$$Y(\mathbf{s}) = \int_{T^d} \exp(i\mathbf{s}'\lambda) dM(\lambda)$$

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

よって表現できる (Gikhman and Skorokhod (1974), Rosenblatt (1985), Yaglom (1987)). ここで $\mathbf{s} = (s_1, \dots, s_d)'$, $\mathbf{h} = (h_1, \dots, h_d)'$, $\lambda = (\lambda_1, \dots, \lambda_d)'$ とする。また $\mathbf{K} = \mathbf{Z}$ のときは $T = [-\pi, \pi]$, $\mathbf{K} = \mathbf{R}$ のときは、 $T = (-\infty, \infty)$ となる。

$M = \{M(\lambda), \lambda \in T^d\}$ は d 次元複素数値直交増分過程、 $F(\lambda)$ は d 次元非負測度で、 $E|M(\Delta)|^2 = F(\Delta) (\Delta \subset T^d)$ および $E(M(\Delta_1)\overline{M(\Delta_2)}) = 0 (\Delta_1, \Delta_2 \subset T^d, \Delta_1 \cap \Delta_2 = \phi)$ を満たす。 $F(\lambda)$ を定常過程と同様にスペクトル分布関数と呼ぶ。 $F(\lambda)$ 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と言う。

6. 定常確率場に対するモデル

本節では定常確率場に対するモデルを考える。時空間定常確率場を叙述するために、我々がモデルに期待するであろう条件として、Ma (2008) は以下の 3 つを挙げている。

- A Simple Stochastic Representation
- A Closed-Form Spectral Density Function
- A Closed-Form of the Covariance Function

さらに加えて Ma (2008) は原文のまま引用すると「..., just as a stationary discrete-time ARMA time series. In general, however, only one or two tractable forms may be available for a spatial or spatio-temporal stationary random field, ...」と主張している。

確かに (3.5), (3.7), (3.8) から分かるように時系列解析における ARMA モデルはこれら 3 条件を兼ね備えている。また彼が主張するように現時点では確率場に対してこれらの条件を満たしているモデルは数少ない。

統計モデルの導入法としては (i) $Y(\mathbf{s})$ 自身に対して直接導入する方法と (ii) 自己共分散関数 (スペクトル密度関数) に対して導入する方法がある。前者については本稿では省略する。興味のある読者は Cressie (1993), Guyon (1995), Gelfand *et al.* (2010) を参照されたい。本稿では 2 番目の方法で導入されたモデルを 2 つ紹介する。一つは等方型 (isotropic) モデルでありもう一つは分離型 (separable) モデルである。

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を等方型モデルと言う。等方型モデルを構成するにはまず 1 変数の正定値 (positive definite) 関数 $C_0(x) (x \in \mathbf{R})$ を用意する。次に x に $\|\mathbf{h}\|$ を代入し

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

を自己共分散関数とすればよい。

しかしここで一つ障害が生じる。 $C_0(x)$ が正定値関数でも、 $d \geq 2$ のとき $C(\mathbf{h})$ は必ずしも正定値関数にならないことである (Christakos (1984), Cressie (1993, p. 84)). 任意の d に対して $C(\mathbf{h})$ が正定値関数になるための必要十分条件は $C_0(x) = \int_0^\infty \exp(-x^2 u^2) dG(u)$ を満たす場合である。ここで $G(x)$ は有界非減少関数とする。またこの必要十分条件は $g(x) = C_0(x^{1/2})$ とおいたとき、 $g(x) (x \geq 0)$ が完全単調 (completely monotone) 関数になることと同値である。 $g(x)$ が $x \geq 0$ において非負の値をとり、任意の n に対して n 次の導関数が $(-1)^n d^n g(x)/dx^n \geq 0 (0 < x < \infty)$ を満たすとき、完全単調関数と言う (Schoenberg (1938), Feller (1971), Cressie (1993), Stein (1999)). 例として $g(x) = \exp(-\alpha x^\gamma) (0 < \alpha, 0 < \gamma \leq 1)$ がある。

$C_0(x)$ に対するモデルとして Matérn 族 (Matérn class) がある (Matérn (1947, 1960, 1986), Handcock and Stein (1993), Stein (1999)). 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 (Stein (1999)). \mathcal{K}_ν は変形された第 2 種ベッセル関数 (Modified Bessel function of the second kind) である。 α, ν, ϕ は正定数で、 ϕ が分散を決める尺度パラメータ、 α が形状を決めるパラメータ、 ν が滑らかさを決めるパラメータで ν が大きいほど、平均 2 乗収束の意味で高次の微分が可能になる。具体的には $\nu = 1/2$ のとき $C_0(x) = \pi \phi \alpha^{-1} \exp(-\alpha|x|)$,

$\nu = 3/2$ のとき $C_0(x) = \frac{1}{2}\pi\phi\alpha^{-3} \exp(-\alpha|x|)(1 + \alpha|x|)$ である. スペクトル密度関数は $\omega = \|\lambda\|$ のみに依存して

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

と言う簡潔な表現を持つ.

ちなみに Guttorp and Gneiting (2006) は Matérn 族にまつわる歴史をコンパクトにまとめており興味深い. それによると Matérn は, 今日 Matérn 族と呼ばれるモデルを 1947 年の博士論文 (Matérn (1947)) において提案し (指導教員は Harald Cramér), それが 1960 年スウェーデン森林研究所の紀要論文 (Matérn (1960)) として流布し, さらに 4 分の 1 世紀以上を経て 1986 年再版されたようである (Matérn (1986)). また物理学者 von Kármán (1948a, b) はある種の風洞実験データの自己共分散関数としては $\nu = 1/3$ が良くフィットすると主張している. 一方 Whittle (1954) は独立に偏微分方程式を解くことにより, $\nu = 1$ となる自己共分散関数を導出している. 既に半世紀以上前に, 時空間統計解析の先駆けとなった碩学達の見聞の明に感服する次第である.

図 1 は分散 σ^2 と α を固定して, $\nu = 0.5, 1.0, 1.5$ と変化させたときの (左上より時計回り) Matérn 族からシミュレーションにより発生させた実現値である. たしかに ν が大きくなるにつれて色の変化が滑らかになって行く. これは左下のグラフのように ν が大きくなるにつれて自己共分散関数の減衰が緩慢になることも平仄の合う結果である.

次に分離型モデル (Separable Model) では自己共分散関数が 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} の成分を二つのグループに分離して

$$C(\mathbf{h}) = C_1(\tilde{\mathbf{h}}_1)C_2(\tilde{\mathbf{h}}_2)$$

$$\tilde{\mathbf{h}}_1 = (h_1, \dots, h_m)', \quad \tilde{\mathbf{h}}_2 = (h_{m+1}, \dots, h_d)'$$

とする. ここで $C_i(\tilde{\mathbf{h}}_i)(i = 1, 2, \dots)$ は各々正定値関数である. 対応してスペクトル密度関数も $C_i(\tilde{\mathbf{h}}_i)(i = 1, 2, \dots)$ のスペクトル密度関数 $f_i(\tilde{\lambda}_i)(i = 1, 2, \dots)$ の積 $f(\lambda) = f_1(\tilde{\lambda}_1)f_2(\tilde{\lambda}_2)$ によって表現される. ここで $\lambda = (\lambda_1, \dots, \lambda_d)', \tilde{\lambda}_1 = (\lambda_1, \dots, \lambda_m)', \tilde{\lambda}_2 = (\lambda_{m+1}, \dots, \lambda_d)'$ とする.

長期記憶定常確率場に対しても等方型モデルおよび分離型モデルをスペクトル密度関数を通して定義できる. たとえば $d = 2$ の場合, 分離型モデルのスペクトル密度関数は

$$f(\lambda) \sim |\lambda_1|^{-d_1}|\lambda_2|^{-d_2} \quad (0 < d_1, d_2 < 1)$$

によって, 等方型モデルのスペクトル密度関数は

$$f(\lambda) \sim (\lambda_1^2 + \lambda_2^2)^{-d} \quad (0 < d < 2)$$

によって各々表現される (Ludeña and Lavielle (1999)).

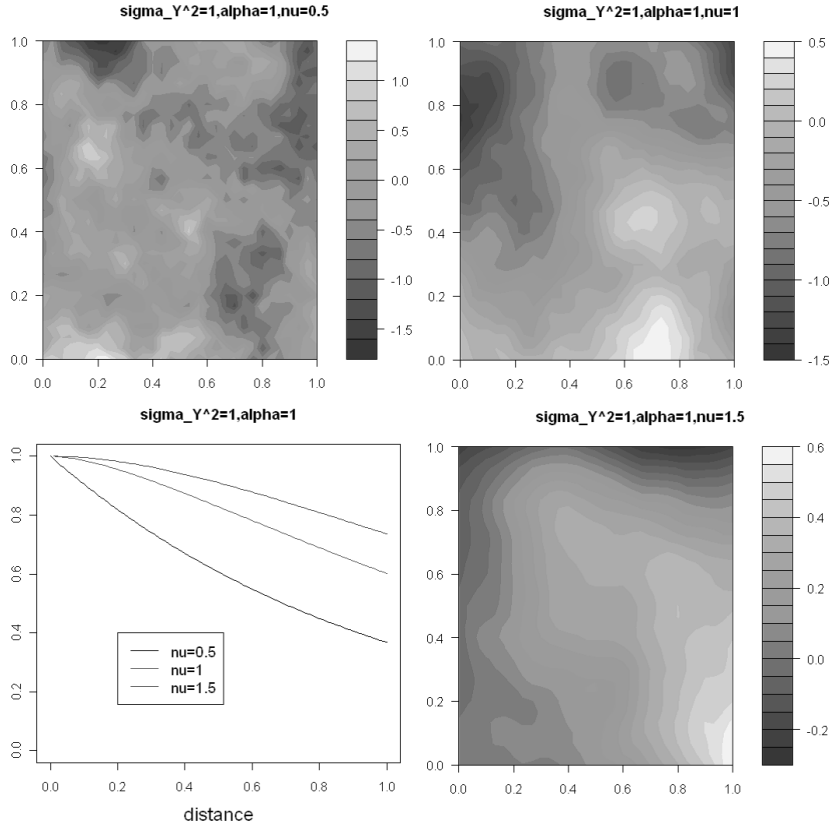


図1 Simulation of Matérn class

なお本節で述べた内容に関し明解でかつ詳しい解説書として間瀬 (2010), Sherman (2011) を挙げておく。

7. 定常確率場のパラメータ推定

本節では4節で説明した Whittle 推定量の定常確率場への一般化について概説する。現在までの主な結果として

- Spatial Short-Memory (Regularly Spaced): Dahlhaus and Künsch (1987)
- Spatial Long-Memory (Regularly Spaced): Ludeña and Lavielle (1999)
- Spatial Short-Memory Gaussian (Irregularly Spaced): Matsuda and Yajima (2009)
- Spatial Long-Memory (Non)Gaussin (Irregularly Spaced): Who ? When?

がある。最後のケースは筆者の知る限りではまだ解かれていないようである。この問題が解決されれば Whittle 推定量の漸近的性質はほぼすべて解明されたことになる。

まず “Regularly Spaced” と言うサンプリング・スキームは時系列データにおいて等間隔で観測される場合の一般化である。等間隔の格子点上でデータが観測され、その領域が拡大していくことを想定している。Increasing Domain Asymptotics とも呼ばれている。具体的には観測領域は直方体 $P_n = \prod_{i=1}^d [1, 2, \dots, n_i]$ 、標本数は $n = \prod_{i=1}^d n_i$ となり、 $n_i \rightarrow \infty (i = 1, 2, \dots, d)$ とする。

Whittle 推定量を構成するには、一見すると $d = 1$ と同じくピリオドグラム

$$I_n(\lambda) = \frac{|\sum_{\mathbf{s} \in P_n} \exp(-i\mathbf{s}'\lambda)Y(\mathbf{s})|^2}{(2\pi)^d n}$$

を用いれば良さそうに見えるが、 $d \geq 2$ では端効果 (edge effect) とする厄介な問題が生じる。自己共分散関数の自然な推定量は、時系列解析と同じく標本自己共分散関数 $\hat{C}(\mathbf{h}) = \sum_{\mathbf{s}, \mathbf{s}+\mathbf{h} \in P_n} Y(\mathbf{s})Y(\mathbf{s}+\mathbf{h})/n$ であるが、バイアスが d とともに増加して

$$E(\hat{C}(\mathbf{h})) - C(\mathbf{h}) = O(n^{-1/d}) \tag{7.1}$$

となる (Guyon (1995)). 従って時系列解析の場合のように $\sqrt{n}(\hat{C}(\mathbf{h}) - C(\mathbf{h}))$ の極限分布を考えると、 $d \geq 2$ の場合は (7.1) のバイアス項が無視できなくなる。原因は観測地点の境界が $d = 1$ (時系列解析) の場合は始点と終点の「点」、 $d = 2$ の場合は「線」、 $d = 3$ の場合は「面」と次元 d の増加とともに増えていくことにある。これを端効果と言う。

この効果を打ち消すために、時系列解析において提案された tapered ピリオドグラムを時空間統計解析に一般化して利用する。標本自己相関関数およびピリオドグラムは $Y(\mathbf{s}) = 0 (\mathbf{s} \notin P_n)$ と見なしていることに等しい。そこで taper と呼ばれる $h(u) : [0, 1] \rightarrow [0, 1]$ かつ $\lim_{u \rightarrow 0, 1} h(u) = 0$ を満たす連続関数 $h(u)$ を用いて、 $Y(\mathbf{s})$ に $\underline{h}(\mathbf{s})Y(\mathbf{s})$ を代入する。ここで $\underline{h}(\mathbf{s}) = \prod_{i=1}^d h(s_i/n_i)$ とする。このようにすれば \mathbf{s} が P_n の境界に近づくとき観測値が緩やかに 0 に収束する。数式で表現すれば tapered ピリオドグラムは

$$I_n^T(\lambda) = \frac{|\sum_{\mathbf{s} \in P_n} \exp(-i\mathbf{s}'\lambda)\underline{h}(\mathbf{s})Y(\mathbf{s})|^2}{(2\pi)^d H_n}$$

によって定義される。ここで $H_n = \sum_{\mathbf{s} \in P_n} \underline{h}^2(\mathbf{s})$ とする。 $h(u) \equiv 1$ のときは $I_n(\lambda)$ に等しい。

このとき \mathbf{Y}_n を定常確率場、 θ をそのパラメータとして

$$L_n^*(\theta; \mathbf{Y}_n) = \int_{(-\pi, \pi]^d} \left(\log f(\lambda; \theta) + \frac{I_n^T(\lambda)}{f(\lambda; \theta)} \right) d\lambda$$

を最小化する推定量を $d \geq 2$ の場合の Whittle 推定量と言う。この推定量は任意の d に対して強一致性を持つが、残念ながら $d \leq 3$ の場合のみ漸近正規性を持つ (Dahlhaus and Künsch (1987)). $d \geq 4$ の場合は依然として端効果が存在し、さらにどのようにして推定量を改善すべきかは未解決の問題である。

次に “Irregularly Spaced” のサンプリング・スキームは色々と考えられるが、ここでは観測地点 \mathbf{s}_i は

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

によって定義する. ここで $\mathbf{u}_i = (u_{i,1}, \dots, u_{i,d})' (i = 1, \dots, n)$ は d 次元の i.i.d. 確率ベクトル列でその密度関数 $g(\mathbf{x})$ は $[0, 1]^d$ に含まれる compact support を持つとする.

$A_j (j = 1, 2, \dots, d)$ を固定して, $n \rightarrow \infty$ とする場合を Infill Asymptotics と言う. ここでは k を引数として, $k \rightarrow \infty$ のとき観測値の総数を増大させ ($n = n_k \rightarrow \infty$) と同時に観測領域も拡大させていく ($A_j(k) \rightarrow \infty (j = 1, \dots, d)$). このサンプリング・スキームを Mixed Domain Asymptotics と言う. このとき離散フーリエ変換 (discrete Fourier transform) およびピリオドグラムを各々

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

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

により定義する. ここで $G = \int_{[0,1]^d} g(\mathbf{s})^2 d\mathbf{s}$, S_k は直方体 $S_k = [0, A_1] \times \dots \times [0, A_d]$, $|S_k|$ はその体積 $|S_k| = A_1 \times \dots \times A_d$ とする. $J_k(\lambda)$ はフーリエ変換 $(2\pi)^{-\frac{d}{2}} |S_k|^{\frac{1}{2}} \int_{S_k} Y(\mathbf{s}) \cdot \exp(-i\mathbf{s}' \lambda) d\mathbf{s}$ を近似する確率変数である.

G は未知なので通常のカーネル関数を用いて

$$\hat{G} = m^{-d} \sum_{i_1=1}^m \dots \sum_{i_d=1}^m \hat{g} \left(\frac{i_1}{m}, \dots, \frac{i_d}{m} \right)^2$$

$$\hat{g}(x_1, \dots, x_d) = \frac{1}{n_k \delta^d} \sum_{j=1}^{n_k} K \left(\frac{u_{j,1} - x_1}{\delta}, \dots, \frac{u_{j,d} - x_d}{\delta} \right)$$

によって推定する. ここで K は \mathbf{R}^d 上のカーネル関数, δ はバンド幅とする. そして

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

を最小にする $\theta(\hat{\theta}_k$ とおく) を, このサンプリング・スキームにおける Whittle 推定量と呼ぶことにする. ここで D は 対称 ($\lambda \in D \rightarrow -\lambda \in D$) なコンパクト集合とする. 注意すべきことはこの場合も端効果によりピリオドグラムに大きなバイアスが生じ, それを補正する $c_k(\theta)$ を $L_k(\theta)$ の中に挿入する必要がある.

ある条件の下で $k \rightarrow \infty$ のとき, 任意の d に対して $\hat{\theta}_k$ は θ_0 に確率収束 (弱一貫性, 強一貫性ではない) する. 一方端効果のため $d \leq 3$ のときのみ漸近正規性が成立して,

$$|S_k|^{1/2} (\hat{\theta}_k - \theta_0) \xrightarrow{L} N(0, 2bW(\theta_0)^{-1})$$

となる (Matsuda and Yajima (2009)). ここで

$$b = \frac{\int_{[0,1]^d} |g(\mathbf{s})|^4 d\mathbf{s}}{\left(\int_{[0,1]^d} g(\mathbf{s})^2 d\mathbf{s}\right)^2}$$

$$W(\theta) = (2\pi)^{-d} \int_D \left(\frac{\partial \log f(\lambda; \theta)}{\partial \theta}\right) \left(\frac{\partial \log f(\lambda; \theta)}{\partial \theta}\right)' d\lambda$$

とする. 興味深いのはシュワルツの不等式より $b \geq 1$ となり, 漸近分散が最小になるのは $g(\mathbf{x}) \equiv 1$ すなわち一様分布の場合である.

8. 今後の課題

最後に本節では今後解決すべき課題のいくつかを列挙する.

(1) 非等方型・非分離型モデルの開発とその推測理論

6節で紹介した等方型モデル・分離型モデルは表現が簡潔でありその結果パラメータ推定もし易いが, 地価・風速などのように距離だけでなく方向に相関が依存するデータ, あるいは天候の変化など時間遅れを伴う空間相関を持つデータの解析に応用することはできない. このようなデータを解析するために非等方型・非分離型を特殊ケースとして含むようなモデルがいくつか提案されている. 大まかに分けると以下の4つのアプローチがある (Ma (2008)).

- 確率偏微分方程式
- スペクトル密度関数
- 自己共分散関数とその線形結合
- 混合法

ここでは紙幅の関係で最初の3つについて概説する.

(a) 確率偏微分方程式

理解の補助としてまず $d = 1$ の場合から始める. この場合は確率微分方程式になるが, 例えば以下の方程式

$$\left(\frac{d}{dt} + \alpha\right) Y(t) = \epsilon(t) \tag{8.1}$$

の解を求めることを考える. ただしこの方程式は形式的な展開であり, $\epsilon(t)$ も直観的に期待値0, 単位時間当たりの分散が σ^2 の連続時間正規「白色雑音」とする. 数学的に厳密な表現は

$$dY(t) = -\alpha Y(t)dt + \sigma dW(t)$$

となる. ここで $\{W(t)\}$ は 3 節で導入したブラウン運動 ($B_{1/2}(t), \sigma = 1$) とする. この解は Ornstein-Uhlenbeck 過程と呼ばれ, 確率積分

$$Y(t) = Y(0) \exp(-\alpha t) + \sigma \int_0^t \exp(-\alpha(t-s)) dW(s)$$

によって表現される (Karatzas and Shreve (1991)). もし $Y(0) \sim N(0, \sigma^2/2\alpha)$ を満たせば, $\{Y(t)\}$ は, 自己共分散関数 $C(h) = (\sigma^2/2\alpha)e^{-\alpha|h|}$, スペクトル密度関数 $f(\lambda) = \sigma^2/(\lambda^2 + \alpha^2)$ の定常過程になる. また離散時点で観測された $\{Y(t)\} (t = 0, 1, 2, \dots)$ は $\phi = e^{-\alpha}$ の AR(1) モデルになる.

次に Whittle (1954) は $d = 2$ の場合について, ラプラス型の偏微分方程式

$$\left(\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2} - \phi^2 \right) Y(s_1, s_2) = \epsilon(s_1, s_2) \quad (8.2)$$

を提案し, 6 節で紹介した自己共分散関数およびスペクトル密度関数が

$$C(h_1, h_2) = \phi \| \mathbf{h} \| \mathcal{K}_1(\phi \| \mathbf{h} \|)$$

$$f(\lambda_1, \lambda) = \frac{\sigma^2}{(\lambda_1^2 + \lambda_2^2 + \phi^2)^2}$$

によって与えられる等方型定常確率場を導いている. ただし (8.2) も (8.1) と同様に形式展開であり, $\epsilon(s_1, s_2)$ も直観的に期待値 0, 単位「面積」当たりの分散が σ^2 の正規「白色雑音」を想定している.

さらに Jones and Zhang (1997) は, 任意の d に対して, 確率偏微分方程式

$$\left[\left(\sum_{i=1}^d \frac{\partial^2}{\partial s_i^2} - \phi^2 \right)^p - c \frac{\partial}{\partial t} \right] Y(\mathbf{s}, t) = \epsilon(\mathbf{s}, t) \quad (8.3)$$

を提案し, 自己共分散関数およびスペクトル密度関数が

$$C(\mathbf{h}, t) = \frac{\| \mathbf{h} \|^{-d/2+1} \sigma^2}{2(2\pi)^{d/2} c} \times \int_0^\infty \frac{u^{d/2} e^{-(u^2 + \phi^2)^p t/c}}{(u^2 + \phi^2)^p} J_{d/2-1}(u \| \mathbf{h} \|) du$$

$$f(\lambda, \tau) = \frac{\sigma^2}{(\| \lambda \|^2 + \phi^2)^{2p} + c^2 \tau^2}$$

によって与えられる定常確率場を導いている. ここで J_ν は第一種 Bessel 関数である.

(8.2), (8.3) の形式展開の数学的正当化は, Ruiz-Medina *et al.* (2004) や Kelbert *et al.* (2005) によってなされている. ただし現時点では筆者自身の力不足により完全には確認していないことを断っておく.

(8.2), (8.3) は右辺の項を確率変数にすることにより確定的な偏微分方程式を「確率化」している. 本来の確定的な偏微分方程式は時間的空間的な熱伝導の変化を記述するなど物理的現象から出発しているが, 確率化によって解いた定常確率場が実際の時空間データに対してどのような含意を持つか必ずしも明らかではないと思われる.

(b) スペクトル密度関数

スペクトル密度関数にモデルを導入する方法として2つ紹介しておく.

一つは Stein (2005) により提案されたモデルで, スペクトル密度関数は

$$f(\lambda, \tau) = \{c_1(a_1^2 + \|\lambda\|^2)^{\alpha_1} + c_2(a_2^2 + \tau^2)^{\alpha_2}\}^{-\beta}, (\lambda, \tau) \in \mathbf{R}^d \times \mathbf{R}$$

によって定義される. ここで $c_1 = \sigma^{-2}$, $c_2 = c^2\sigma^{-2}$, $a_2 = 0$, $\alpha_1 = 2p$, $\alpha_2 = 1$, $\beta = 1$ を代入すると, 前述の Jones and Zhang (1997) が提案したモデルに一致する.

二つめは Fuentes *et al.* (2008) によって提案されたモデルで, スペクトル密度関数は

$$f(\lambda, \tau) = \gamma(\alpha^2\beta^2 + \beta^2 \|\lambda\|^2 + \alpha^2\tau^2 + \epsilon \|\lambda\|^2 \tau^2)^{-\nu}, (\lambda, \tau) \in \mathbf{R}^d \times \mathbf{R}$$

によって定義される. ϵ が「非等方性」と「非分離性」の程度を示すパラメータである. 例えば $\epsilon = 1$ とおけば分離型モデル

$$f(\lambda, \tau) = \gamma(\alpha^2 + \|\lambda\|^2)^{-\nu}(\beta^2 + \tau^2)^{-\nu}$$

になる. 一方 $\epsilon = 0$ のときは, スペクトル密度関数は

$$f(\lambda, \tau) = \gamma(\alpha^2\beta^2 + \beta^2 \|\lambda\|^2 + \alpha^2\tau^2)^{-\nu}$$

となり, Matérn 族を一般化したモデルである.

(c) 自己共分散関数

自己共分散関数に対するモデルとして, Gneiting (2002) は

$$C(\mathbf{h}, t) = \frac{\sigma^2}{\psi(t^2)^{d/2}} \phi\left(\frac{\|\mathbf{h}\|^2}{\psi(t^2)}\right), (\mathbf{h}, t) \in \mathbf{R}^d \times \mathbf{R}$$

を提案している. ここで $\phi(t)(t \geq 0)$ は完全単調関数, $\psi(t)(t \geq 0)$ は正値を取る関数でその微分が完全単調関数とする.

以上さまざまなモデルが提案され, 百家争鳴の感がある. ちなみに Stein (2005) は “a laundry list of potential models” と呼んでいる. それぞれのモデルが持つ含意, モデル・フィッティングとモデル比較の方法, モデル適合度の評価方法などまだまだやるべきことが残されている.

(2) 非定常モデルの開発

時系列解析と同じく時空間統計解析においてもグローバルな範囲で観測されるデータに対して定常性を仮定するのは現実性を欠く場合が多い. さまざまな非定常確率場が提案されているが, ここでは2つ紹介する.

(a) 固有定常確率場 (intrinsic stationary random field)

任意の \mathbf{h} に対して以下の 2 条件が成立する確率場 $\{Y(\mathbf{s})\}$

$$E(Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})) = 0$$

$$E[Y(\mathbf{s} + \mathbf{h}) - Y(\mathbf{s})]^2 = 2\gamma(\mathbf{h})$$

を固有定常確率場と言う。 $2\gamma(\mathbf{h})$ をバリオグラム (variogram), $\gamma(\mathbf{h})$ をセミバリオグラム (semivariogram) と呼ぶ。時系列解析における単位根モデルを時空間データへ一般化したモデルと見なせる。

例えば 3 節で説明した FBM モデルを多次元 (multidimensional) FBM モデルに拡張することも可能である。多重確率積分で定義する二つの方法がある。一つは

$$Y(\mathbf{s}) = \int_{\mathbf{R}^d} \frac{\exp(i\mathbf{s}'\lambda) - 1}{\|\lambda\|^{d/2+H}} S(\lambda/\|\lambda\|) dW(\lambda)$$

によって与えられる。ここで $W(\lambda)$ は 5 節の $M(\lambda)$ と同じく直交増分を持ち $dE|W(\lambda)|^2 = d\lambda$ を満たす正規過程とする。 S が非等方性を表す関数である (Istas (2007))。このときバリオグラムは

$$2\gamma(\mathbf{h}) = \int_{\mathbf{R}^d} \frac{|\exp(i\mathbf{h}'\lambda) - 1|^2}{\|\lambda\|^{d+2H}} S^2(\lambda/\|\lambda\|) d\lambda$$

となる。 $S(\lambda/\|\lambda\|) \equiv 1$ のときは、等方的なモデルとなり

$$2\gamma(\mathbf{h}) = C \|\mathbf{h}\|^{2H}$$

を満たす。ここで C は正定数とする。従って 1 次元 FBM モデルの多次元への拡張になっている (Samorodnitsky and Taqqu (1994))。

二つめは

$$Y(\mathbf{s}) = \int_{\mathbf{R}^d} \frac{\exp(i\mathbf{s}'\lambda) - 1}{\|\lambda\|^{d/2+H(\lambda)}} dW(\lambda)$$

によって定義される。 $H(\lambda)$ は 0 次同次すなわち任意の定数 $c(\neq 0)$ に対して $H(c\lambda) = H(\lambda)$ が成立する関数である (Bonami and Estrade (2003))。このときバリオグラムは

$$2\gamma(\mathbf{h}) = \int_{\mathbf{R}^d} \frac{|\exp(i\mathbf{h}'\lambda) - 1|^2}{\|\lambda\|^{d+2H(\lambda)}} d\lambda$$

となる。長期記憶性の程度を示す H がベクトル λ の方向に依存するモデルである。

以上 2 つのモデルの相違を明らかにすること、 $S(\lambda/\|\lambda\|)$ あるいは $H(\lambda)$ の推定方法、実際データへの応用可能性などは今後の課題である。

(b) 進化的スペクトル関数

このモデルは時系列解析において Priestley (1965) が提案した進化的 (evolutionary) スペクトル密度関数の時空間統計解析版で

$$Y(\mathbf{s}, t) = \int_{\mathbf{R}^d} \int_{\mathbf{R}} \exp(i\mathbf{s}'\lambda + i\tau t) \phi_{\mathbf{s},t}(\lambda, \tau) dM(\lambda, \tau)$$

$$E|dM(\lambda, \tau)|^2 = d\lambda d\tau$$

によって定義される。ここで $\phi_{\mathbf{s},t}(\lambda, \tau)$ が時空間に依存した振幅を表す関数である (Fuentes *et al.* (2008)).

このとき共分散関数 $Cov(Y(\mathbf{s}_1, t_1), Y(\mathbf{s}_2, t_2))$ は

$$C(\mathbf{s}_1, t_1; \mathbf{s}_2, t_2) = \int_{\mathbf{R}^d} \int_{\mathbf{R}} \exp(i(\mathbf{s}_1 - \mathbf{s}_2)'\lambda + i(t_1 - t_2)\tau) \times \phi_{\mathbf{s}_1, t_1}(\lambda, \tau) \bar{\phi}_{\mathbf{s}_2, t_2}(\lambda, \tau) d\lambda d\tau$$

となる。

$\phi_{\mathbf{s},t}(\lambda, \tau)$ に具体的な関数を与えることによりさまざまなモデルを構築できる。例えば $\phi_{\mathbf{s},t}(\lambda, \tau) = \phi_{\mathbf{s}}^{(1)}(\lambda) \phi_t^{(2)}(\tau)$ のような分離型の関数を代入すると、共分散関数

$$Cov(Y(\mathbf{s}_1, t_1), Y(\mathbf{s}_2, t_2)) = C^{(1)}(\mathbf{s}_1, \mathbf{s}_2) C^{(2)}(t_1, t_2)$$

を持つ、非定常分離型確率場になる。

あるいは K 個の地点 $\mathbf{s}_j (j = 1, \dots, K)$ をあらかじめ選び、 $\phi_{\mathbf{s}_j}(\lambda, \tau)$ をその近傍での時空間的構造を説明する関数とする。地点 \mathbf{s} が \mathbf{s}_j が遠ざかるにつれ影響が小さくなるようにカーネル関数 $K(\mathbf{s} - \mathbf{s}_j)$ を用いてウエイトをつけ、

$$\phi_{\mathbf{s},t}(\lambda, \tau) = \sum_{j=1}^k K(\mathbf{s} - \mathbf{s}_j) \phi_{\mathbf{s}_j}(\lambda, \tau)$$

とする確率場も構築できる。

\mathbf{s}_j の選択方法、 $\phi_{\mathbf{s},t}(\lambda, \tau)$ の推定方法あるいは分離型のような特定の関数形に対する検定方法などはまだ完全には確立されていない。

(3) 大規模データの処理 n 個の地点 $\mathbf{s}_i (i = 1, \dots, n)$ でデータ $Y(\mathbf{s}_i)$ が観測され、その共分散行列を

$$C = (c_{ij}), \quad n \times n, \quad c_{ij} = Cov(Y(\mathbf{s}_i), Y(\mathbf{s}_j))$$

とする。パラメータの最尤推定量の導出や未知のデータに対する最良不偏線形予測量 (Best Linear Unbiased Predictor (BLUP)), 時空間統計ではクリギング (Kriging) とも呼ばれる (間瀬 (2010)) の構成には、逆行列 C^{-1} の計算が必要となる。しかし操作回数は n^3 に比例

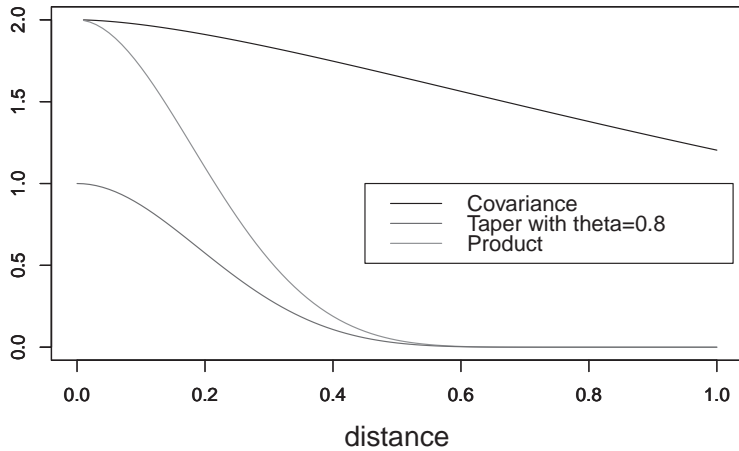


図2 Overview of Covariance Tapering

する。リモートセンシングのように瞬時に大量のデータが採取できる場合、實際上計算不可能なオーダーになる。例えば $n = 1,000$ であれば 10 億に上る。

これを回避する一つの方法として covariance tapering がある。説明のためにここでは定常確率場を考える。真の自己共分散関数を $C(\mathbf{h})$ としたとき、これにサポートがコンパクトな自己相関関数 $C_\theta(\mathbf{h})$ (共分散関数ではない。すなわち $C_\theta(\mathbf{0}) = 1$) をかけた

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

をあたかも真の自己共分散関数と見なし、それから構成される共分散行列の逆行列を計算する方法である。例えばある正定数 θ に対して

$$C_\theta(\mathbf{h}) = 0, \quad \|\mathbf{h}\| > \theta$$

が成立すると仮定する。図示すると図 2 ($\theta = 0.8$) のようになり、一番上の自己共分散関数を中央の自己共分散関数により近似する。

ここで $n \times n$ 行列 $\Sigma_{tap} = (\sigma_{ij,tap})(n \times n)$ を

$$\sigma_{ij,tap} = C_{tap}(\mathbf{s}_i - \mathbf{s}_j)$$

によって定義すれば、

$$\sigma_{ij,tap} = 0, \quad \|\mathbf{s}_i - \mathbf{s}_j\| > \theta$$

が成立する。従って Σ_{tap} は 0 となる成分が多い “sparse” な行列となり C に比べはるかに高速に逆行列を計算できる。

しかしその理論的正当性が証明されているのは、推定および予測とも Matérn 族のみのようである (Du *et al.* (2009), Furrer *et al.* (2006), Kaufman *et al.* (2008), Wang and Loh (2011)). 他の時空間モデルにおいても正当化できるかは未解明である。また covariance tapering は一定の距離以上離れた観測値は無相関と見なすことに等しい。時系列解析で言えば、一定の時間差以上離れた観測値は無相関と見なす MA モデルを当てはめていることに相当する。一方時系列解析における AR モデルのように時空間統計解析においてもマルコフ型のモデルを仮定して推定や予測を行うことも可能である。この方法と covariance tapering との優劣も今後明らかにすべきことである。

(4) その他

冒頭に述べたようにこの分野は広範囲に渡っている。本稿で概説したトピックはほんの「氷山の一角」であり、「その他」と一括りにするが、今後さらに発展を期すべき分野は多々ある。例えば 感染性疾病の特定地域への集積、動植物の植生・生態の変化、さらに付け加えるまでもなく 2011 年 3 月 11 日に発生した東日本大震災などの発生メカニズム、これらの時空間的構造をモデル化する時空間点過程がある。あるいは都市・産業の集積、欧州統合など経済活動の国際化が各国の経済成長に与える影響など、経済データの時空間的変動や特性を解明する分野として時空間計量経済学がある。これらの参考文献としては Arbia (2006), Arbia and Baltagi (2009), Cressie (1993), Gelfand *et al.* (2010), LeSage and Pace (2009) がある。

謝辞

本稿は科学研究費補助金基盤研究 (A)No. 19200020 の補助を部分的に受けて執筆された。図 1, 2 は東京大学大学院経済学研究科平野敏弘氏が作成したものである。氏のご厚意により掲載させていただいた。また初稿を改訂するに当たり日本統計学会和文誌編集委員長青嶋誠氏および査読者を務められた編集委員の方からは大変有益なコメントをいただいた。併せて謝意を表したい。

参 考 文 献

- Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, Wiley.
- Arbia, G. (2006). *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*, Springer.
- Arbia, G. and Baltagi, B. H. (2009). *Spatial Econometrics: Methods and Applications*, Springer.
- Banerjee, A., Dolad, J., Galbraith, J. W. and Hendry, D. F. (1993). *Co-integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press.
- Beran, J. (1992). *Statistics for Long-Memory Processes*, Chapman and Hall.
- Bonami, A. and Estrade, A. (2003). Anisotropic analysis of some Gaussian models, *J. Fourier Anal. Appl.*, **9**, 215–236.
- Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control* (1st. ed.), Holden Day.

- Brockwell, P. J. and Davis, R. A. (1991). *Time Series : Theory and Methods* (2nd ed.), Springer.
- Christakos, G. (1984). On the problem of permissible covariance and variogram models, *Water Resour. Res.*, **20**, 251–265.
- Cressie, N. (1993). *Statistics for Spatial Data* (rev. ed.), Wiley.
- Dahlhaus, R. and Künsch, H. R. (1987). Edge effects and efficient parameter estimation on stationary random fields, *Biometrika*, **74**, 877–882.
- Doukhan, P., Oppenheim, G. and Taqqu, M. (2002). *Theory and Applications of Long-Range Dependence*, Birkhäuser.
- Du, J., Zhang, H. and Mandrekar, V. (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators, *Ann. Statist.*, **37**, 3330–3361.
- Dunsmuir, W. (1979). A central limit theorem for parameter estimation in stationary vector time series and its application to model for a signal observed with noise, *Ann. Statist.*, **7**, 490–506.
- Dunsmuir, W. and Robinson, P. M. (1981). Parametric estimators for stationary time series with missing observations, *Adv. Appl. Probab.*, **13**, 129–146.
- Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer.
- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, vol. II (2nd ed.), Wiley.
- Finkenstädt, B., Held, L. and Isham, V. (2007). *Statistical Methods for Spatio-Temporal Systems*, Chapman & Hall/CRC.
- Fox, R. and Taqqu, M. S. (1985). Noncentral limit theorems for quadratic forms in random variables having long-range dependence, *Ann. Probab.*, **13**, 428–446.
- Fox, R. and Taqqu, M. S. (1986). Large-sample properties of parameter estimation for strongly dependent stationary Gaussian time series, *Ann. Statist.*, **14**, 517–532.
- Fox, R. and Taqqu, M. S. (1987). Central limit theorems for quadratic forms in random variables having long-range dependence, *Probab. Th. Rel. Fields*, **74**, 213–240.
- Fuentes, M., Chen, L. and Davis, J. M. (2008). A class of nonseparable and nonstationary spatial temporal covariance functions, *Environmetrics*, **19**, 487–507.
- Fuller, W. A. (1976). *Introduction to Statistical Time Series* (1st ed.), Wiley.
- Furrer, R., Genton, M. G. and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets, *J. Comp. Graph. Statist.*, **15**, 502–523.
- Gelfand, A. E., Diggle, P. J., Fuentes, M. and Guttorp, P. (2010). *Handbook of Spatial Statistics*, Chapman & Hall/CRC.
- Gikhman, I. I. and Skorokhod, A. V. (1974). *The Theory of Stochastic Processes I*, Springer.
- Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate, *Prob. Th. Rel. Fields*, **86**, 87–104.
- Gneiting, T. (2002). Nonseparable stationary covariance functions for space-time data, *J. Amer. Statist. Assoc.*, **97**, 590–600.
- Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-range time series models and fractional differencing, *J. Time Ser. Anal.*, **1**, 15–29.
- Grenander, U. and Szegő, G. (1984). *Toeplitz Forms and their Applications* (2nd ed.), Chelsea.
- Guoriéroux, C. (1997). *ARCH Models and Finance Applications*, Springer.
- 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
- Handcock, M. S. and Stein, M. L. (1993). A Bayesian analysis kriging, *Technometrics*, **35**, 403–410.
- Hannan, E. J. (1970). *Multiple Time Series*, Wiley.
- Hannan, E. J. (1973). The asymptotic theory of linear time series models, *J. Appl. Probab.*, **10**, 130–145.
- Hosking, J. R. M. (1981). Fractional differencing, *Biometrika*, **68**, 165–176.
- Hosoya, Y. (1997). A limit theory for long-range dependence and statistical inference on related models, *Ann. Statist.*, **25**, 105–137.

- Hosoya, Y. and Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes, *Ann. Statist.*, **10**, 132–153, Correction: (1993). **21**, 1115–1117.
- Istas, J. (2007). Identifying the anisotropical function of a d -dimensional Gaussian self-similar process with stationary increments, *Statist. Inf. Stoch. Proc.*, **10**, 97–106.
- Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models* (2nd ed.), Oxford University Press.
- Jones, R. H. and Zhang, Y. (1997). Models for continuous stationary space-time processes, *Modelling Longitudinal and Spatially Correlated Data* (eds. T. G. Gregoire *et al.*), Lecture Notes in Statistics 122, Springer, 289–298.
- Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus* (2nd ed.), Springer.
- 刘屋武昭, 矢島美寛, 田中勝人, 竹内啓 (2003). 経済時系列の統計 その数理的基礎, 統計科学のフロンティア 8, 岩波書店.
- Kaufman, C., Schervish, M. and Nychka, D. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets, *J. Amer. Statist. Assoc.*, **103**, 1545–1555.
- Kelbert, M., Leonenko, N. and Ruiz-Medina, M. D. (2005). Fractional random fields associated with stochastic fractional heat equations, *Add. Appl. Probab.*, **37**, 108–133.
- Lavancier, F. and Philippe, A. (2011). Some convergence results on quadratic forms for random fields and application to empirical covariances, *Prob. Th. Rel. Fields*, **149**, 493–514.
- LeSage, J. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*, Chapman & Hall/CRC.
- Ludeña, C. and Lavielle, M. (1999). The Whittle estimator for strongly dependent stationary Gaussian fields, *Scand. J. Statist.*, **26**, 433–450.
- Ma, C. (2008). Recent developments on the construction of spatio-temporal covariance models, *Envir. Res. and Risk Asses.*, **22**, S39–S47.
- Manderbrot, B. B. and van Ness, J. W. (1968). Fractional Brownian motions, fractional noise and applications, *SIAM Rev.*, **10**, 422–437.
- 間瀬茂 (2010). 地球統計学とクリギング法 R と GeoR によるデータ解析, オーム社.
- Matérn, B. (1947). *Metoder att Uppskatta Noggranheten vid Linje- och Provytetaxering*, Stockholm: Medd. Staten 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.), Springer.
- Matsuda, Y. and Yajima, Y. (2009). Fourier analysis of irregularly spaced data on R^d , *J. Roy. Statist. Soc.*, **B71**, 191–217.
- Priestley, M. B. (1965). Evolutionary spectral and non-stationary processes, *J. Roy. Statist. Soc.*, **B27**, 204–237.
- Priestley, M. B. (1981). *Spectrum Analysis and Time Series*, Vols. 1 and 2, Academic Press.
- Robinson, P. M. (2003). *Time Series with Long Memory*, Oxford University Press.
- Rosenblatt, M. (1985). *Stationary Sequences and Random Fields*, Birkhäuser
- Ruiz-Medina, M. D., Anguko, J. M. and Anh, V. V. (2004). Fractional random fields on domain with fractal boundary conditions, *Inf. Dim. Anal. Quantum Probab. Rel. Top.*, **7**, 395–417.
- Samorodnitsky, G. and Taqqu, M. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman & Hall/CRC.
- Schoenberg, I. J. (1938). Metric spaces and completely monotone functions, *Ann. Math.*, **39**, 811–841.
- Shaman, P. (1976). Approximations for stationary covariance matrices and their inverses with application to ARMA models, *Ann. Statist.*, **4**, 292–301.
- Sherman, M. (2011). *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties*, Wiley.
- Sinai, Ya. G. (1976). Self-similar probability distribution, *Th. Probab. Appl.*, **21**, 64–80.
- Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*, Springer.
- Stein, M. L. (2005). Space-time covariance functions, *J. Amer. Statist. Assoc.*, **100**, 310–321.

- Taylor, S. J. (1994). Modelling stochastic volatility; A review and comparative studies, *Mathematical Finance*, **4**, 183-204.
- Tong, H. (1990). *Non-Linear Time Series: A Dynamic System Approach*, Oxford University Press.
- von Kármán, T. (1948a). Progress in the statistical theory of turbulence, *J. Marine Res.*, **7**, 252-264.
- von Kármán, T. (1948b). Progress in the statistical theory of turbulence, *Proc. Nat. Acad. Sci.*, **34**, 530-539.
- Walker, A. M. (1964). Asymptotic properties of least-square estimates of parameters of the spectrum of a stationary nondeterministic time series, *J. Austral. Math. Soc.*, **4**, 363-384.
- Wang, D. and Loh, W.-L. (2011). On fixed-domain asymptotics and covariance tapering in Gaussian random fields, *Electronic J. Statistics*, **5**, 238-269.
- Whittle, P. (1954). On stationary processes in the plane, *Biometrika*, **41**, 434-449.
- Whittle, P. (1961). Gaussian estimation in stationary time series, *Bull. Internat. Statist. Inst.*, **39**, 105-129.
- Yaglom, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions*, vol. I, Springer.
- Yajima, Y. (1985). On estimation of long-memory time series models, *Austral. J. Statist.*, **27**, 303-320.
- Yajima, Y. (1992). Asymptotic properties of estimates in incorrect ARMA models for long-memory time series, *New Directions in Time Series Analysis, Vol. II* (eds. D. Brillinger *et al.*), 375-382.