時空間計量経済学
その現況と今後の展望
於.日本経済学会春季大会2017

矢島美寛 (yajima@e.u-tokyo.ac.jp)

東北大学大学院経済学研究科

内容

- 定義
- 歴史
- データ
- モデル
- 今後の展望

定義（空間計量経済学）

Definition

"Models and theoretical instruments of spatial statistics and spatial data analysis to analyse various economic effects such as externalities, interactions, spatial concentration and many others."(Spatial Econometrics Association(2006 発足) の規約より)

歴史

Reference: The Lecture Note of Prof. Giuseppe Arbia(Universita Cattolica del Sacro Cuore,Rome) Sept. 2016 at University of Tokyo

"Spatial Econometrics " was coined by Paelinck, J.H. and Klaasen,L.H.(1979).

A Short History of Spatial Statistics and Econometrics

- Moran(1950). *J.Roy.Statist.Soc.*
- Whittle(1954). *Biometrika*
- Cliff and Ord(1973). *Spatial Autocorrelation*
- Besag(1974). *J.Roy.Statist.Soc.*
- Bennett(1979). *Space-Time Series*

歴史(続き)

- Anselin(1988). *Spatial Econometrics*
- Cressie(1993). *Statistics for Spatial Data*
- Arbia(2006). *Spatial Econometircs*
- Cressie and Wikle(2011). *Statistics for Spatio-Temporal Data*
- Arbia(2014). *Spatial Econometircs*

歴史(続き)

90’s: An important event in spatial econometircs!(cited from Arbia(2016))

- *”New Economic Geography”*(Fujita, Kurgman and Venables(1999)
- A theoretical structure that justifies a spatial analysis for economic data to face topics like regional convergence and spatial concentration of economic activities
- Many new N.E.G. theories lead to the specification of models that are susceptible of empirical validation and require spatial econometrics tools(e.g. β)-convergence, Barro and Sala-i-Martin(1995))

歴史(続き)

さらに 20 世紀の統計学の花形の 1 つ時系列解析から 21 世紀は時空間統計解析の時代!?

理由

- 様々な地球規模の問題 (経済・伝染性の疾病 (pandemic)・環境問題 (温暖化・オゾン層の破壊) など)
- データ収集のためのインフラストラクチャー・科学技術の進歩 (リモートセンシング・GPS・GIS など)
- 解析手段の進歩 (高速大容量計算機・統計解析ソフトウエア (R など))

歴史(続き)

26th May 2006:Birth of the Spatial Econometircs Association(SEA)

- **Founded by:** Anselin,Arbia,Baltagi,Kelejian,Paelinck,Prucha,Robinson
- **The annual conferences** have been held every year since then in U.S.A. and Europe. The conferece of this year was held for the first time in Singapore, Asia two weeks ago.
- **Special issues:**Empirical Economics 2007,$\cdots$,Economic Modelling2010,$\cdots$,Geographical Analysis 2012,$\cdots$,Annals of Regional Science 2016.

データ

数学的表現

- パラメータ:$\boldsymbol{R}$ あるいは $\boldsymbol{Z} = \{0, \pm 1, \pm 2, \ldots\}$ の d 次元直積空間 (集合). 統一的には $\boldsymbol{K}^d$
- 観測地点 (site):$\boldsymbol{s}(\in \boldsymbol{K}^d)$
- 観測データ:$Y(\boldsymbol{s})$(一変量データ, 多変量データ)
- 例 $d = 2$:$\boldsymbol{s}$ は緯度, 経度.$Y(\boldsymbol{s})$ はその地点の地価.
- $d = 3$:$\boldsymbol{s}$ は緯度, 経度, 高さ.$Y(\boldsymbol{s})$ はその地点における気温.
- $D(\subset \boldsymbol{K}^d)$:$\boldsymbol{s}$ の動く領域
- 確率場 (random field):$\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D\}$(データの全体)

データ(続)

数学的表現（続）

- 観測時点も考慮する場合:$d = 4$ とする. あるいは時点を強調し $Y(s,t)$ と表す

領域に基づく時空間データの3分類

(1) 地点参照データ (point-referenced data)

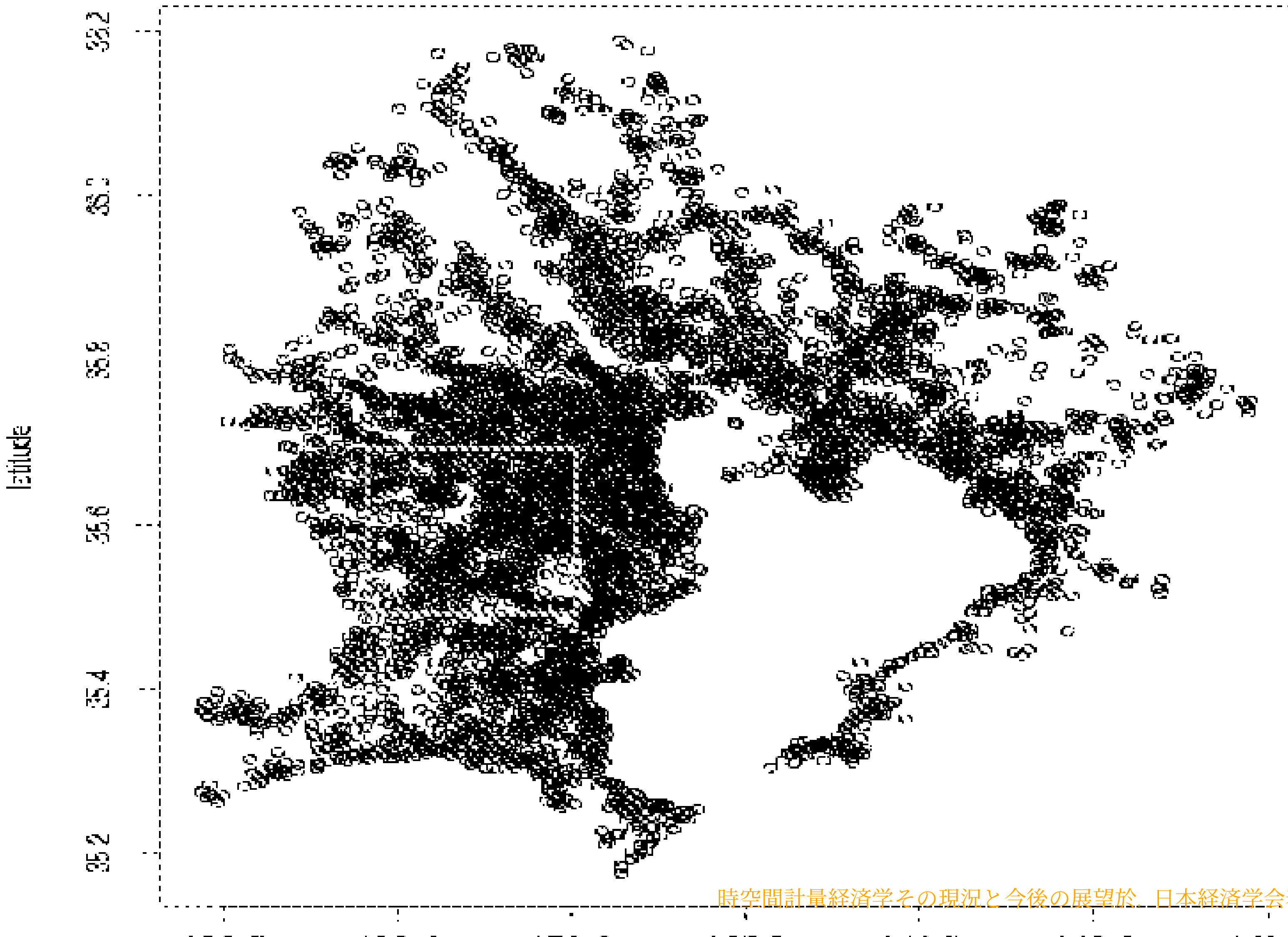
- D:正の体積を持つ d 次元直方体を含む $\boldsymbol{R}^d$ の部分集合
- s は D 上を連続的に変化
- 例:気温, 風速, 風向データなど.

領域に基づく時空間データの3分類

(2) 格子データ (lattice data) あるいは地域データ (areal data)

- D:高々可算個の点からなる $\boldsymbol{R}^d$ の部分集合
- 観測地点の間隔は規則的な場合と不規則な場合がある
- 規則的な場合は D を $\boldsymbol{Z}^d$ あるいはその部分集合で表す.
  - 例:各格子に画素が与えられた画像データ
- 不規則な場合の例:2001 年の首都圏公示地価観測地点.
- 行政地区における集計データなどではその地区の中心都市のデータとして割り当てることもある
  - 例. 都道府県別の失業率 $\rightarrow$ 都庁, 道庁, 府庁, 県庁所在地に割り当てる

関東地方公示地価観測地点



領域に基づく時空間データの3分類

(3) 点配置データ (point pattern data)

- D そのものが確率変数
  - 例. ある事象が生起した地点のデータ解析
  - 地点 s で地震が起きたときには $W(s) = 1$. 地震の震度を $Y(s)$
  - 起きなかったときには $W(s) = 0$.
  - $D = \{s|W(s) = 1\}$:地震の起きた地点の全体
- 領域 D および観測値 $Y(s)$ の特性について解析
- 起きた時点まで考慮すれば $D = \{(s,t)|W(s,t) = 1\}$

モデル

時系列解析と時空間統計解析**時系列解析**

- One-parameter Stochastic Process
  $\{X_t | t \in \boldsymbol{R} = (-\infty, \infty)\}$
- 自然な順序の導入 (時間順序. 過去-現在-未来)
- 時間相関、時間変動のモデル化

時空間統計解析

- Multi-parameter Stochastic Process(Random Field)
  $\{X_t | t \in \boldsymbol{R}^d\} (2 \le d)$
- 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？

モデル(続)

Tobler's first law of geography

"*Everything is related to everything else, but close things are more related than things that are far apart.*"

定常確率場(地点参照データ)

(弱) 定常確率場の定義: $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \boldsymbol{K}^d\}$

(i) 期待値は $\boldsymbol{s}$ に依存せず一定 $E(Y(\boldsymbol{s})) = \mu$. 以下では簡単のため $\mu = 0$ とする.

(ii) 共分散はベクトル差 $\boldsymbol{t} - \boldsymbol{s}$ のみに依存し,

$C(\boldsymbol{t} - \boldsymbol{s}) = E(Y(\boldsymbol{t})Y(\boldsymbol{s})),\ \boldsymbol{t}, \boldsymbol{s} \in \boldsymbol{K}^d$.

注意 $d = 1$ の場合は時系列解析における弱定常過程.

$\{C(\boldsymbol{h}),\ \boldsymbol{h} = \boldsymbol{t} - \boldsymbol{s} \in \boldsymbol{K}^d\}$ を弱定常過程の場合と同様に自己共分散関数 (autocovariance function) と呼ぶ.

定常確率場(続)

スペクトル表現

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

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

$\boldsymbol{s} = (s_1, \ldots, s_d)'$, $\boldsymbol{h} = (h_1, \ldots, h_d)'$, $\lambda = (\lambda_1, \ldots, \lambda_d)'$, $'$ は転値ベクトル.

$$\boldsymbol{K} = \boldsymbol{Z} \to T = (-\pi, \pi]$$

$$\boldsymbol{K} = \boldsymbol{R} \to T = (-\infty, \infty)$$

定常確率場(続)

スペクトル表現(続き)

- $M = \{M(\lambda), \lambda \in T^d\}$:$d$ 次元複素直交増分過程
- $F(\lambda)$:d 次元非負測度でスペクトル分布関数
- $E|M(\Delta)|^2 = F(\Delta)(\Delta \subset T^d)$
- $E(M(\Delta_1)\overline{M(\Delta_2)}) = 0(\Delta_1,\ \Delta_2(\Delta_1 \cap \Delta_2 = \phi) \subset T^d)$
- 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と言う

定常確率場に対するモデル

Desirable Properties

- A Simple Stochastic Representaiton
- A Closed-Form Spectral Density
- A Closed-Form Covariance Function

Just as an ARMA model in Time Series! However, only one or two tractable forms may be available(for a spatial or spatio-temporal stationary random field!
Ma(2008)*Stoch.Envir.Res.Risk Assess.*)

定常確率場に対するモデル(続)

統計モデルの導入法

方法 1. $Y(s)$ 自身に対して直接導入

方法 2. 自己共分散関数 (スペクトル密度関数) に対して導入

自己共分散関数に対するモデル

(1) 等方型 (Isotropic Model) モデル

自己共分散関数 $C(\boldsymbol{h})$ が距離 $\| \boldsymbol{h} \| = \sqrt{\sum_{i=1}^{d} h_i^2}$ のみに依存,
方向には無関係.

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

$$C_0(x)(x \in \boldsymbol{R}) : \text{One Variable Positive Definite Function}$$

注意 $C_0(x)$p.d function $\rightarrow C(\boldsymbol{h})$p.d function$)(d \geq 2)$ は必ずしも言えない！

自己共分散関数に対するモデル

$$
\begin{aligned}
& C(\boldsymbol{h}) : \text{p.d. function } \forall d \\
\Leftrightarrow\quad & C_0(x) = \int_0^\infty \exp(-x^2u^2)dG(u) \\
& G(u) : \text{bounded nondecreasing function} \\
\Leftrightarrow\quad & f(x)(= C_0(x^{1/2}) : \text{completely monotone}
\end{aligned}
$$

Remark $f(x)$ is completely monotone for $x \geq 0$ if $(-1)^n d^n f(x)/dx^n \geq 0$ for $0 < x < \infty$

Example. $f(x) = \exp(-\alpha x^\gamma),\ \alpha > 0, 0 < \gamma \leq 1.$

Matérn 族 ($C_0(x)$ に対するモデル)

一般形 (Matérn(1980)2nd ed. Springer)

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

α, ν, ϕ : positive constants

$\mathcal{K}_{\nu}$: Modified Bessel function

ϕ : scale parameter

α : shape and convergence parameters

ν : smoothness parameter

Matérn 族の例

Guttorp and Gneiting(2006)*Biometrika*

$$\nu = 1/2 \to C_0(x) = \pi\phi\alpha^{-1}\exp(-\alpha|x|)$$

$$\nu = 3/2 \to C_0(x) = \frac{1}{2}\pi\phi\alpha^{-3}\exp(-\alpha|x|)(1+\alpha|x|)$$

$$\nu = 1/3 \text{ von Karman}(1881-1963) \text{ turbulence}$$

$$\nu = 1 \text{ Whittle}(1954) Biometrika$$

スペクトル密度関数は $\omega = \| \lambda \|$ のみに依存

$$f(\omega) \sim \phi(\alpha^2 + \omega^2)^{-\nu - d/2}$$

自己共分散関数に対するモデル

(2) 分離型モデル (Separable Model)

$C(\boldsymbol{h})$ を 2 個以上の正定値関数の積として表現. 例えば

$$C(\boldsymbol{h}) = C_1(\tilde{\boldsymbol{h}}_1)C_2(\tilde{\boldsymbol{h}}_2)$$
$$\tilde{\boldsymbol{h}}_1 = (h_1, \ldots, h_m)',\ \tilde{\boldsymbol{h}}_2 = (h_{m+1}, \ldots, h_d)'$$

スペクトル密度関数

$$f(\lambda) = f_1(\tilde{\lambda}_1)f_2(\tilde{\lambda}_2)$$
$$\lambda = (\lambda_1, \ldots, \lambda_d)'$$
$$\tilde{\lambda}_1 = (\lambda_1, \ldots, \lambda_m)',\ \tilde{\lambda}_2 = (\lambda_{m+1}, \ldots, \lambda_d)'$$
$$f_1(\tilde{\lambda}_1),\ f_2(\tilde{\lambda}_2) : \text{nonnegative functions}$$

SARAR モデル(格子・地域データ)

$$
\begin{aligned}
Y(\boldsymbol{s}_i) &= \lambda \sum_{j=1}^{n} w_{ij} Y(\boldsymbol{s}_j) + \sum_{k=1}^{K} X(\boldsymbol{s}_i, k)\beta_k + U(\boldsymbol{s}_i),\ i = 1, \ldots, n \\
U(\boldsymbol{s}_i) &= \rho \sum_{j=1}^{n} w_{ij} U(\boldsymbol{s}_j) + \epsilon(\boldsymbol{s}_i),\ i = 1, \ldots, n
\end{aligned}
$$

Matrix Representation

$$
\begin{aligned}
\boldsymbol{Y} &= \lambda \boldsymbol{W} \boldsymbol{Y} + \boldsymbol{\beta} \boldsymbol{X} + \boldsymbol{U} \\
\boldsymbol{U} &= \rho \boldsymbol{W} \boldsymbol{U} + \boldsymbol{\epsilon}
\end{aligned}
$$

SARAR モデル(続)

ここで $\boldsymbol{X}$ は説明変数, $\boldsymbol{\epsilon} = (\epsilon(\boldsymbol{s}_1), \epsilon(\boldsymbol{s}_2) \ldots, \epsilon(\boldsymbol{s}_n))'$ は n 次元正規分布 $N(\mathbf{0}, \sigma_\epsilon^2 I)$ に従い, $\boldsymbol{W} = (w_{ij})$ は隣接行列 (connectivity matrix) といい, 対角成分は $w_{ii} = 0 (i = 1, \ldots, n)$ とする.

SARAR モデル(続)

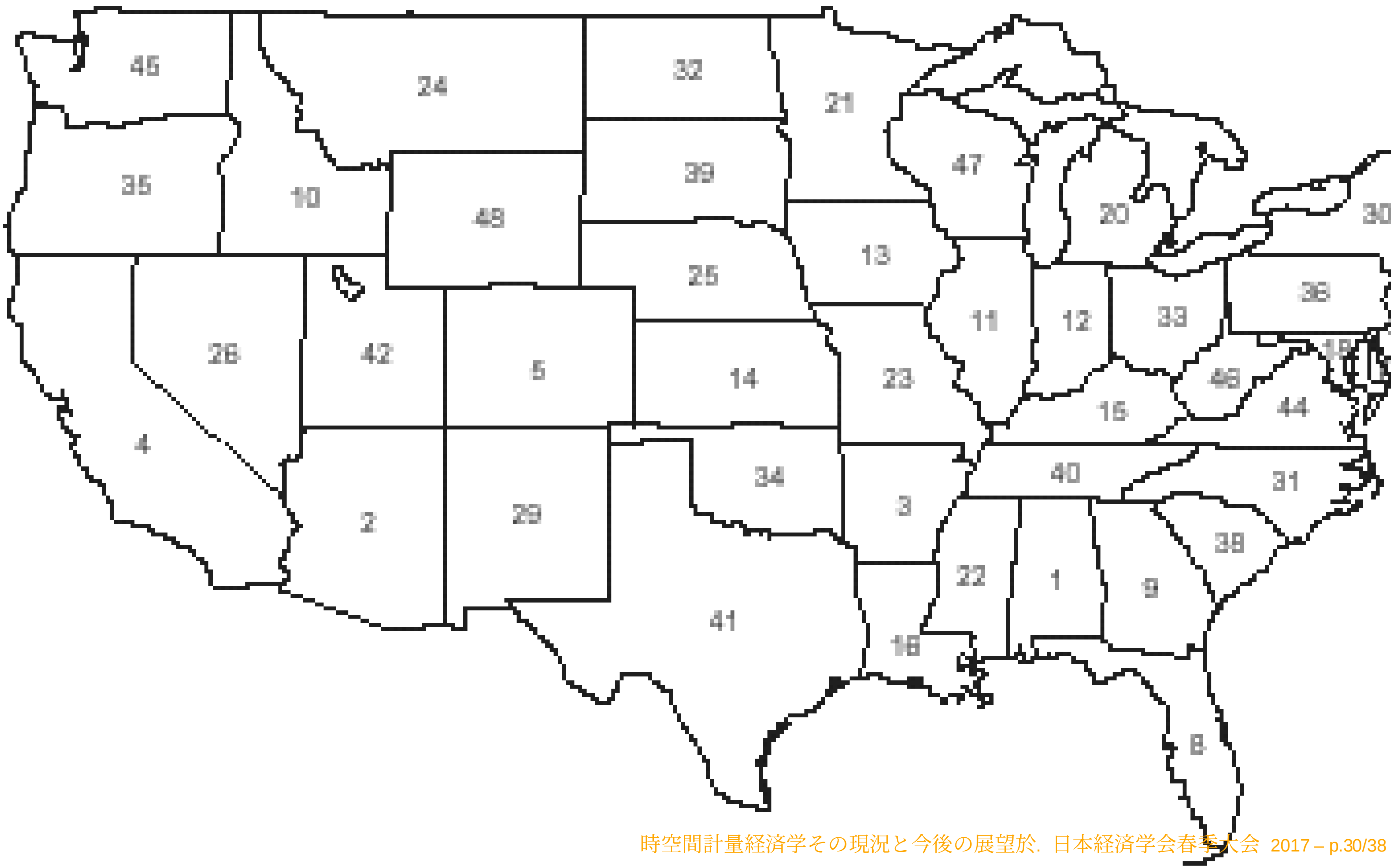
隣接行列の例

$$w_{ij} = \begin{cases} 1, & \boldsymbol{s}_i \text{ and } \boldsymbol{s}_j \text{ are adjacent } (i \neq j) \\ 0, & i = j \\ 0, & \text{otherwise} \end{cases}$$

例. アメリカ合衆国の州. 番号は州名をアルファベット順に並べたときの順番を意味している. Alabama 州 (番号 1. 以下同様) と隣接する州は, Florida(8), Georgia(9), Mississippi(22), Tennessee(40) の 4 州である.

$w_{1,8} = w_{1,9} = w_{1,22} = w_{1,40} = 1,\ w_{ij} = 0,$ otherwise.

SARAR モデル(続)


45
24
32
21
35
10
48
39
47
20
30
13
25
36
11
12
33
26
42
5
14
23
46
44
4
15
34
40
31
2
29
3
38
22
1
9
41
16
8

SARARモデル(続)

SARA includes 5 models.

- $\beta_k = 0$ and $\lambda = 0$ or $\rho = 0$:Pure SAR model
- $\lambda = 0$ and $\rho = 0$:Lagged dependent variables model
- $\lambda \neq 0$ and $\rho = 0$:Spatial Lag Model(SLM)
- $\lambda = 0$ and $\rho \neq 0$:Spatial Error Model(SEM)
- $\lambda \neq 0$ and $\rho \neq 0$:SARAR(1,1)-The complete model

今後の課題

- Nonseparable and Anisotropic Model の開発と推測理論
  例. 時間遅れを伴う空間相関関数のモデル化
- 非定常モデルの開発と推測理論
- SARAR モデルにおける隣接行列の選択
  例. 時間距離, 文化的・社会的類似性の数量化
- Many other problems

本日の講演

- 松田　新たな非定常確率場のモデル
- 堤　　自然科学・人文社会科学における実証分析 (R の応用例など)
- 中島　空間集積に対する実証分析 (Modifirable Areal Unit Problem の克服など)

参考文献

Anselin, L.(1988). *Spatial Econometrics* Kluwer, Boston.

Arbia,G.(2006). *Spatial Econometrics:Statistical Foundations and Applications to Regional Convergence*. Springer.

Arbia,G.(2014). *A Premier for Spatial Econometrics:With Application in R.* Palgrave Macmillan. (空間計量経済学入門. 2016 年 勁草書房. 堤盛人監訳).

Barro,R.J. and Sala-i-Martin,X.(1995). *Economic Growth*. McGraw-Hill.

Bennet,R.J.(1979). *Spatial Time-Series Analysis*. Pion, London.

Besag,J.(1974).Spatial interaction and the statistical analysis of lattice systems(with discussion). *J.Roy. Statist. Soc. Ser. B* **36** 192-236.

参考文献（続）

Cliff,A.D. and Ord, J.K.(1973). *Spatial Autocorrelation*.Pion, London.

Cressie, N.(1993).*Statistics for Spatial Data*. rev. ed. Wiley.

Cressie, N. and Wikle,C.K.(2011). *Statistics for Spatio-Temporalal Data*. rev. ed. Wiley.

Fujita, M., Krugman,P. and Venables,A.J.(1999). *The Spatial Economy*. The MIT Press, Cambridge.

Moran,P.A.P(1950).The interpretation of statistical maps. *J.Roy. Statist. Soc. Ser. B* **10** 243-251.

Paelinck, J.H. and Klaasen,L.H.(1979). *Spatial Econometrics*. Gower, Westmead, Farnborough.

Whittle,P.(1954).On stationary processes in the plane. *Biometrika* **41** 434-449.

参考文献（続）

清水邦夫 (2002). 地球環境データー衛星リモートセンシング. データサイエンス・シリーズ 8 共立出版.

瀬谷創・堤盛人 (2014). 空間統計学. 自然科学から人文・社会科学まで. 朝倉書店.

丹後俊郎・横山徹爾・高橋邦彦 (2007). 空間疫学への招待. 疾病地図と疾病集積性を中心として. 医学統計シリーズ 7 朝倉書店.

間瀬茂・武田純 (2001). 空間データモデリングー空間統計学の応用. データサイエンス・シリーズ 7 共立出版.

矢島美寛 (2008). 時空間統計解析の理論と応用.(21 世紀の統計科学 II. 自然・生物・健康の統計科学. 国友直人・山本拓監修. 国友直人・小西貞則編所収). 東京大学出版会

参考文献（続）

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 巻. 219-244.

予告

2017 年度統計関連学会連合大会

2017 年 9 月 4 日 (月)〜6 日 (水) 於. 南山大学

企画セッション"History and Recent Developments in Spatio-Temporal Statistics"

講演者

Giuseppe Arbia(The invited speaker)

堤盛人

松田安昌

矢島美寛