

第 10 章 空間計量経済モデル


本章では、点オブジェクトや面オブジェクト用いた空間計量経済モデルを取り上げる。第 5 章で紹介したように、空間データの属性値は空間的な自己相関をもつ場合がある。本章では、変数や誤差項に関する空間的な自己相関や系列相関を明示的に取り入れた回帰モデル群を主に扱う。具体的には、**同時自己回帰モデル**、**条件付き自己回帰モデル**、**誤差項の空間的自己回帰モデル**、**空間的自己回帰モデル**、**空間ダービンモデル**、**空間的自己相関を考慮した条件付き回帰モデル**等が挙げられる。また、地区（地点）間の異質性を表現したモデルに、**地理的加重回帰モデル**がある。本章では、これらのモデルを紹介するだけでなく、**一般化線形回帰モデル**や**一般化加法モデル**、**マルチレベルモデル**を用いた地理的特性の表現方法や、空間計量経済モデルのベイズ推定についても扱う。

空間計量経済モデルは、空間疫学分野では標準化死亡比などの分析、都市経済学分野では地価などの不動産市場の空間分析、計量政治学分野では選挙行動の空間分析、生態学分野では生物の個体群・群集の分析等に用いられている。

空間計量経済学全般を学習する上で参考となる最近の文献として、Anselin et al. eds. (2004)[1]、Arbia (2006) [2]、LeSage and Pace (2009)[3]などが挙げられる。地理的加重回帰モデルに焦点を当てた文献として、Fotheringham et al. (2002) [4]などがある。また LeSage (1999)[5]では、ベイズ空間計量経済学について紹介されている¹。

10.1 回帰モデルと空間的自己相関

10.1.1 最小二乗法による回帰モデルの推定

空間データを用いて、回帰モデルを推定するとき、どのような点に留意すべきだろうか。ここでは、具体的に関東圏（一都五県）の市区町村別地価データ（住宅地標準地地価の平均価格、万円/m²）、夜間人口密度（千人/m²）、第三次産業従業員人口密度（千人/m²）などのデータを用いて、住宅地地価を予測する回帰モデルを例に挙げて説明しよう。このうち、地価データの分布を  10.1 に示す。

地区（市区町村） $i(=1, \dots, N)$ の地価 y_i 、夜間人口密度 x_{1i} 、第三次産業従業員人口密

¹ なお本章で示しているベイズ空間計量経済モデルの推定手順については、筆者のホームページ記載のコードを参照されたい。

度 x_{2i} データを用いて、次式のように線形回帰モデルを定式化する。

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

ここで β_0 、 β_1 、 β_2 は回帰係数、 σ^2 は誤差項 ε の分散を意味する。

より一般的に、次式のように書くことにする。

$$y = X\beta + \varepsilon$$

$$\varepsilon \sim N(0, \Sigma)$$

ここで、 y は被説明変数ベクトル、 X は説明変数行列、 β は回帰係数ベクトル、 $\Sigma = \sigma^2 I$ (I は $N \times N$ 単位行列) である。

通常、最小二乗法 (OLS) を用いて回帰モデルを推定した結果は、表 10.1 のようになる。また、通常最小二乗法による地価モデルについて、誤差の空間分布を示すと図 10.2 のようになる。

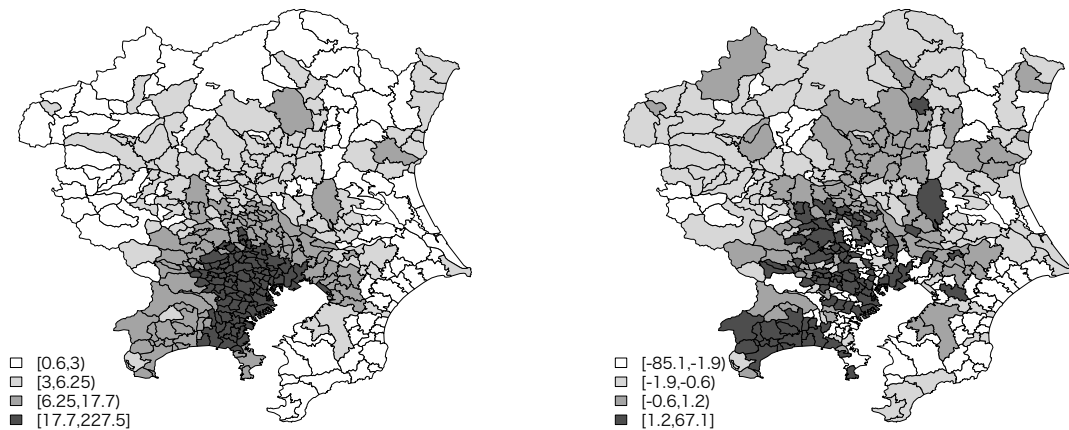


図 10.1 市区町村別地価データ分布 図 10.2 地価モデル (OLS) の誤差項分布

表 10.1 線形回帰モデルの推定結果 (最小二乗法)

変数	回帰係数	t 値
定数項	2.55	4.50
夜間人口密度	1.68	15.85
第三次産業従業人口密度	2.25	28.65
自由度修正済み R^2 値	0.828	

R 分析例

`spdep` パッケージを呼び出し、地価データを属性にもつ関東地方のポリゴンデータ `kanto.shp` を読み込み、`lm()` 関数を用いて線形回帰モデルを推定する。ポリゴンデータの属性テーブルには、地価 (`LPH`)、夜間人口密度 (`POPD`)、第三次産業人口密度 (`EMP3D`) に加え、市区町村代表地点の座標 (`Easting` と `Northing`) や市区町村コード (`JCODE`) などが含まれている。主題図を作成するために、`classInt` パッケージも呼び出す。

```
# パッケージの呼び出し
library(spdep)
library(classInt)
# ポリゴンデータの読み込み
kanto <- readShapePoly("kanto_area.shp", IDvar="JCODE")
# 地価に関する主題図の作成
# 色パレットの作成
pall <- gray.colors(n=4, start=1, end=0.3)
q_kanto <- classIntervals(round(kanto$LPH, 1), n=4,
style="quantile")
q_kanto_Col <- findColours(q_kanto, pall)
plot(kanto, col=q_kanto_Col)
legend("bottomleft", fill=attr(q_kanto_Col, "palette"),
legend=names(attr(q_kanto_Col, "table")), cex=1.2, bty="n")
# GIS データの属性からモデル推定
lph.lm <- lm(LPH~POPD+EMP3D, data=kanto)
summary(lph.lm)
# 地価モデル (lph.lm) の誤差に関する主題図の作成
# 誤差の算出
lph.lm.resid <- resid(lph.lm)
kanto$lm.resid <- lph.lm.resid
# 誤差の表示
# 色パレットの作成
```

```

pall <- gray.colors(n=4,start=1,end=0.3)
q_lm.resid <- classIntervals(round(kanto$lm.resid,1), n=4,
style="quantile")
q_lm.resid_Col <- findColours(q_lm.resid,pall)
plot(kanto,col=q_lm.resid_Col)
legend("bottomleft",fill=attr(q_lm.resid_Col,"palette"),
legend=names(attr(q_lm.resid_Col,"table")), cex=1.2, bty="n")

```

10.1.2 線形回帰モデルのベイズ推定

以下のような事前情報を与えることにより、前出の線形回帰モデルをベイズ推定することができる。ここでは、回帰係数 β の事前情報として平均 0、分散 1×10^{-6} の正規分布 $N(0, 1 \times 10^{-6})$ を与えている。また、 IG は逆ガンマ関数である。

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$

$$\varepsilon_i \sim N(\mu_i, \sigma^2)$$

$$\beta_0 \sim N(0, 1 \times 10^{-6})$$

$$\beta_1 \sim N(0, 1 \times 10^{-6})$$

$$\beta_2 \sim N(0, 1 \times 10^{-6})$$

$$\sigma^2 \sim IG(0.001, 0.001)$$

線形回帰モデルのベイズ推定結果を表 10.2 及び図 10.3 に示す。ここでは、MCMC 期間を 100,000 回、稼働検査（バーンイン）期間を 10,000 回としている。

表 10.2 線形回帰モデルのベイズ推定結果

変数	平均	標準偏差	[2.5%, 97.5%]
定数項	2.55	0.57	[1.42, 3.67]
夜間人口密度	1.68	0.11	[1.47, 1.89]
第三次産業従業人口密度	2.25	0.079	[2.09, 2.40]
σ^2	67.41	5.07	[58.14, 78.01]

R 分析例

`MCMCpack` パッケージの `MCMCregress()` 関数を用いて線形回帰モデルのベイズ推定を行う例を示す。すでに `kanto` データが読み込まれているものとする。

```

# パッケージの呼び出し
library(MCMCpack)
# モデルの推定
lph.mcmc <- MCMCregress(LPH~POPD+EMP3D,data=kanto,
b0=0, B0=1e-6, c0=1e-2, d0=1e-2, mcmc=100000, burnin=10000)
summary(lph.mcmc)
plot(lph.mcmc)

```

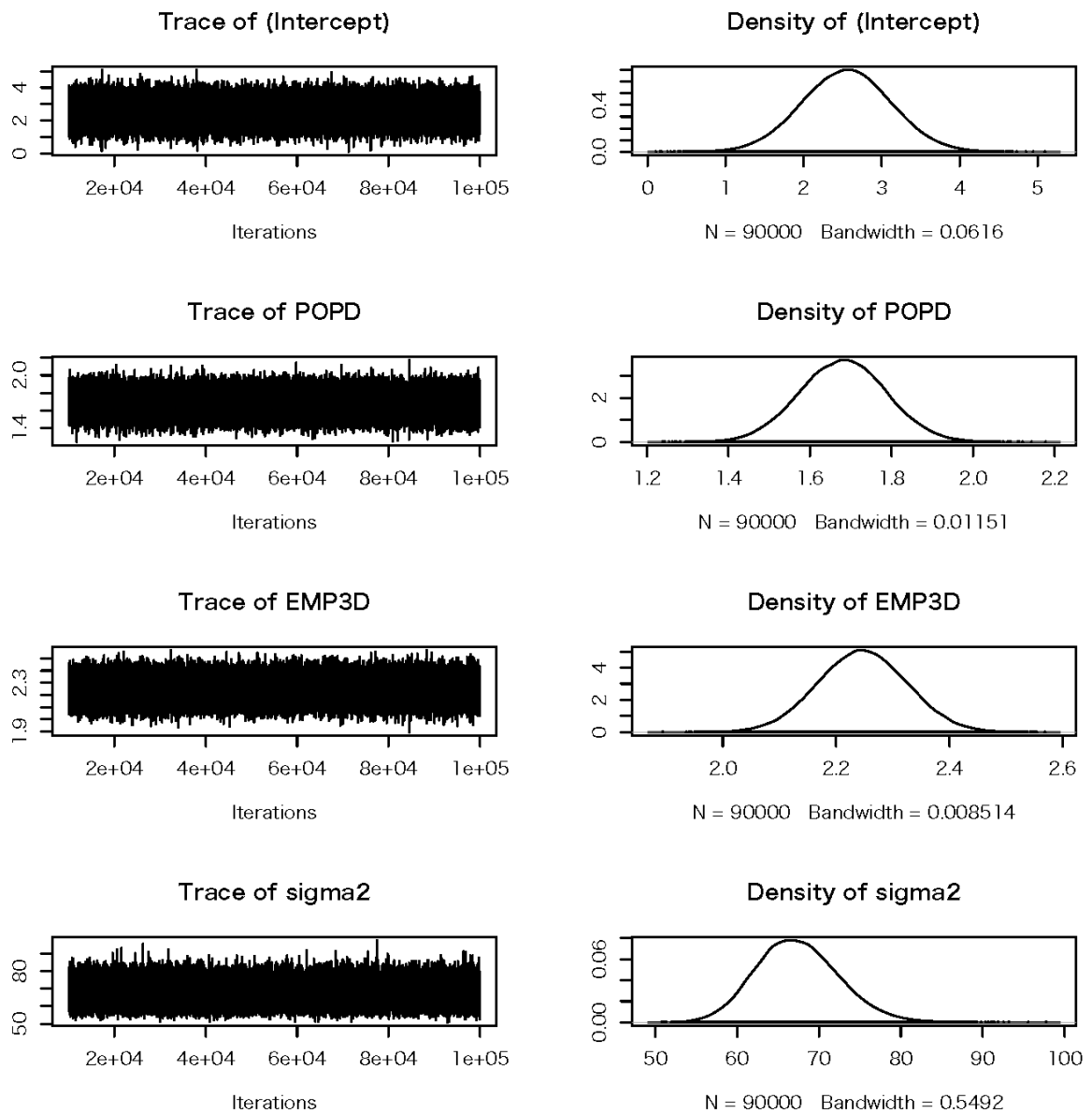


図 10.3 線形回帰モデルのベイズ推定結果

10.1.3 空間的従属性と空間的異質性

ところで、この回帰モデルを推定する際に、本来考慮すべきこととして、まず、観測データや誤差項の**空間的従属性** (spatial dependency) の問題が挙げられる。第5章で紹介したように、Moran's Iなどの指標を用いれば、各変数の空間的自己相関の有無を把握することが出来る。

市区町村代表点の座標を元にドロネー三角網図による空間重み付け行列 W を作成し、地価、夜間人口密度、第三次産業従業人口密度について Moran's I を計算すると、それぞれ 0.76、0.86、0.63 となり、強い空間的自己相関を示している。

被説明変数 y について空間的自己相関が認められる場合には、隣接する地域への空間的波及効果を持つと考えられる。例えば、都市開発などにおける地価上昇や人口減少による地価下落といった問題では、ある地域での地価上昇（下落）が周辺地域の地価上昇（下落）に影響を与えることが知られているが、これは空間的波及効果の一例である。

被説明変数だけでなく、説明変数 X や誤差項 ε に空間的自己相関があると考えられる場合にも、注意が必要である。最尤推定法を用いて回帰モデルを推定するとき、観測データの間で空間的な自己相関があると考えられる場合には、単純な対数尤度関数の和の最大化問題として、パラメータ推定を行うことは出来ない。しかし、空間的自己相関を明示的に取り込むことで最尤推定法を適用できる。

誤差項 ε に空間的自己相関が存在すると考えられる場合、推定されたパラメータが統計的に有意であっても、それは見かけ上の相関にすぎず、特に通常最小二乗法を用いて推定されたパラメータは、一致性も不偏性ももたない。このような場合には、誤差項 ε の地域的なトレンドは、そもそも排除されるべきノイズとして認識できる。この点については、10.3 節で具体的に検討することとしよう。

次に、誤差項 ε や回帰係数 β に関する**空間的異質性** (spatial heterogeneity) の問題が挙げられる。通常の線形回帰モデルでは、誤差項 ε が独立で同一な正規分布に従うと仮定している。このとき、誤差項 ε の**分散均一** (homoscedasticity) $\varepsilon_i \sim N(0, \sigma^2)$ を前提としている。つまり $\sigma_i^2 = \sigma^2$ と仮定している。

しかし、空間データはしばしば地域的・地理的な特徴を有するため、誤差項 ε の分散が地域ごとに異なることがある。すなわち、

$$\varepsilon_i \sim N(0, \sigma_i^2)$$

或いは、

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \sigma_N^2 \end{pmatrix}$$

となることがある。このような場合、誤差項 ε の分散不均一 (heteroschedasticity) を前提としたモデル推定を行うことが要求される。誤差項の分散が均一であるかどうかは、モデルの分散共分散構造を考える上で重要である。

観測データ間の関係、特に説明変数 X と被説明変数 y との関係が地域ごとに異なるような場合には、回帰係数 β の空間的異質性も考慮すべきである。回帰係数 β が対象地域全体で均一であるという仮定をはずして、地域ごとに回帰係数 β を推定するようなモデル (マルチレベルモデルなど) を適用するのが望ましい。また、ベイズ推定法を適用することにより、空間的異質性を考慮した柔軟なモデリングが可能となる。

10.2 可変集計単位問題

3.1 で示した可変集計単位問題は、空間データモデリングにおいても重要な問題である。例えば、メッシュデータを用いて分析する際には、メッシュ規模によってモデルパラメータの推定結果が大きく異なることがある。集計単位の規模を小さくすることにより、政策課題を空間的にきめ細かく検討できると期待できるかもしれない。しかしながら、局地的な空間的自己相関の発生や、小地域ではデータが観測されない、個人情報保護の観点からデータが公開されないといった、小地域統計独自の課題も有する。こうした問題を集計単位の**規模の問題**という。

集計規模が同じであっても、地区の形状によって観測データの集計値が変化することがある。この問題を集計単位の**ゾーニングの問題**という。行政境界を用いて分析を行う際には、規模の問題とゾーニングの問題が同時に発生するため、分析単位を適切に設定することが求められる。

異なった集計地区単位による回帰係数の挙動の安定性を検証するには、モンテカルロシミュレーションなどにより回帰係数の平均と分散を求めるとよい。また、回帰係数の差の検定を行うことで、集計単位に応じた回帰係数の差が統計的に有意かどうかを結論づけるのが望ましい。

10.3 一般化回帰モデル

10.3.1 一般化線形モデル

空間データのモデリングにおいては、集計単位の規模や形状による影響を考慮して、観測データを面積で基準化した変数を用いることがある。人口密度や単位面積あたり地価は、面積で基準化された変数であると言える。説明変数をすべて面積で基準化するような場合には、回帰係数を持たないオフセット項として面積を用いた一般化線形モデルを適用する方法も提案されている。例えば上述の地価モデルの場合であれば、次式のような一般化線形モデルとして定式化できる。ただし S_i は地区 i の面積である。

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \text{offset}(\ln(S_i)) + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

10.3.2 一般化加法モデル

一般化加法モデルは、説明変数の変換を行う関数を組み込んだモデルである。空間計量経済学では、次式のように緯度経度データをもちいた平滑化関数を組み込んだモデルを用いることがある。ただし、*Northing* は経度、*Easting* は緯度である。

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 s(\text{Easting}, \text{Northing}) + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2)$$

表 10.3 に、一般化線形回帰モデルと一般化加法モデルの推定結果の例を示す。

表 10.3 一般化モデルの推定結果

変数	一般化線形モデル		一般化加法モデル	
	回帰係数	t 値	回帰係数	t 値
定数項	-1.29	-2.27	8.57	8.55
夜間人口密度	1.76	16.49	0.21	0.91
第三次産業従業人口密度	2.26	28.66	2.02	23.90
$\text{offset}(\ln(S_i))$	1.00	∞	-	-
$s(\text{Easting}, \text{Northing})$	-	-	25.27	2.21
AIC	2558.0		2531.4	

10.4 自己回帰モデル

空間的な系列相関を回帰モデルに取り入れる方法がいくつか提案されているが、本節では誤差項に空間的な系列相関を明示的に考慮した回帰モデルのうち、同時自己回帰モデル (simultaneous autoregressive model) と条件付き自己回帰モデル (conditional autoregressive model) を紹介する。

10.4.1 同時自己回帰モデル

同時自己回帰モデル (SAR) は、ある地域の誤差項が、自地域を含む他の地域の誤差項と相関関係を取り入れたモデルである。地価モデルを例に挙げると、次式のようなモデルとなる。

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$
$$\varepsilon_i = \sum_{j=1}^N b_{ij} \varepsilon_j + e_i$$

ここで、 b は未知パラメータである。 ε_i は空間的な系列相関を持つ誤差項、 e_i は空間的な系列相関を持たない誤差項を意味する。 b が空間重み付け行列の要素と未知パラメータ λ であらわされる ($b = \lambda w_{ij}$) 場合、10.5.2 で紹介する誤差項の空間的自己相関モデルとなる。モデルの推定結果を表 10.5 に、誤差項の分布を [図 10.4](#) に、それぞれ示す。

10.4.2 条件付き自己回帰モデル

条件付き自己回帰モデルは、その誤差項 ε_i が地域 i を除く誤差項に条件付けられることを明示したモデルである。地域 i 周辺の誤差項を選び $\varepsilon_{j \sim i}$ とすると、その条件付き分布は次式のように表すことができる。

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$$
$$\varepsilon_i | \varepsilon_{j \sim i} \sim N \left(\frac{\sum_{j \sim i} c_{ij} \varepsilon_j}{\sum_{j \sim i} c_{ij}}, \frac{\sigma_{\varepsilon_i}^2}{\sum_{j \sim i} c_{ij}} \right)$$

ここで、 c_{ij} は未知パラメータである。 c_{ij} を空間重み付け行列の要素と未知パラメータ λ であらわす ($c_{ij} = \lambda w_{ij}$) ことができるが、CARモデルでは周辺の全地域の誤差項を用いて条件付けない。

CARモデルは、分散共分散行列 V が以下になるという点で、SARモデルとは異なる。

$$V = (I - \rho W)^{-1} \Sigma$$

誤差項が分散均一の時は、次式のように表される。

$$V = \sigma^2(I - \rho W)^{-1}$$

CAR モデルの推定結果を表 10.4 に、誤差項の空間分布を図 10.5 に、それぞれ示す。SAR モデルと SAR モデルの推定結果から、定数項や POPD、EMP3D に関する回帰係数に大きな差はないが、誤差項の系列相関を示す未知パラメータ λ は 2 倍近い差があることが示された。

表 10.4 自己回帰モデルと条件付き自己回帰モデルの最尤推定結果

変数	SAR		CAR	
	回帰係数	Z 値	回帰係数	Z 値
定数項	3.00	4.26	3.23	4.45
夜間人口密度 (POPD)	1.58	12.73	1.54	12.19
第三次産業従業人口密度 (EMP3D)	2.18	25.31	2.15	24.82
λ	0.22	4.54 ¹⁾	0.44	4.53 ¹⁾
AIC	2552.1		2552.1	

1) Z 値ではなく LR 検定統計量

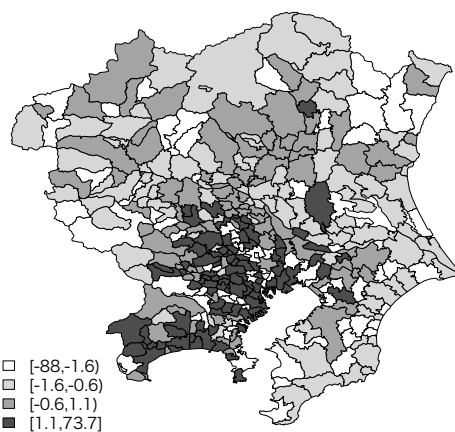
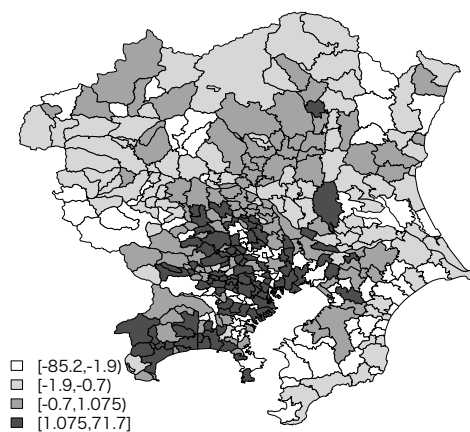


図 10.4 SAR モデルの誤差項分布

図 10.5 CAR モデルの誤差項分布

R 分析例

`spdep()` パッケージの `spautolm()` 関数を用いて、SAR モデルと CAR モデル

を推定する。10.1.1 の分析例の手順でパッケージとデータが用意されているものとする。誤差項の空間的な系列相関を考慮するために、隣接行列を定義した後、`spautolm()` 関数の中で空間重み付け行列を指定する。空間隣接行列は `tri2nb()` 関数を用いてドロネー三角網により作成し、`nb2listw()` 関数により空間重み付け行列を定義している。

```
# 隣接行列の作成
coords <- matrix(0, nrow=length(kanto$LPH), ncol=2)
coords[,1] <- kanto$Easting
coords[,2] <- kanto$Northing
lph.tri.nb <- tri2nb(coords)
# 同時自己回帰 (SAR) モデル
lph.sar <- spautolm(LPH~POPD+EMP3D, data=kanto,
nb2listw(lph.tri.nb, style="W"), family="SAR")
summary(lph.sar)
# 条件付き自己回帰 (CAR) モデル
lph.car <- spautolm(LPH~POPD+EMP3D, data=kanto,
nb2listw(lph.tri.nb, style="W"), family="CAR")
summary(lph.car)
```

10.5 空間的自己相関モデル

10.5.1 空間的自己回帰モデル

空間的自己回帰モデル（空間同時自己回帰モデル、`spatial auto-regression model`）は、被説明変数に空間的従属性を表現したモデルである。このモデルは、空間的波及効果を定式化したモデルであると言える。

$$y = \rho W y + X \beta + \varepsilon$$
$$\varepsilon \sim N(0, \sigma^2 I)$$

分散共分散行列 V は、次式のようになる。

$$V = (I - \rho W)^{-1} \Sigma (I - \rho W)'$$

誤差項の分散が均一 $\sigma_i^2 = \sigma^2$ であるとき、次のように置き換えることができる。

$$V = \sigma^2 \left[(I - \rho W)' (I - \rho W) \right]^{-1}$$

ここで、地価データ y と、空間的従属性を考慮した地価データ W_y との関係性を散布図で示すと、[図 10.6](#) のようになる。[図 10.7](#) と [図 10.8](#) は、それぞれ y と W_y の経験累積分布を示したものである。 y と W_y には相関関係があるように見える。

空間的自己回帰モデルを最尤推定した結果を[表 10.5](#) に、誤差項分布を[図 10.9](#) にそれぞれ示す。線形回帰モデルは $AIC=2554.6$ であり、空間的自己回帰モデルの方が AIC が小さく、当てはまりがよいことがわかる。

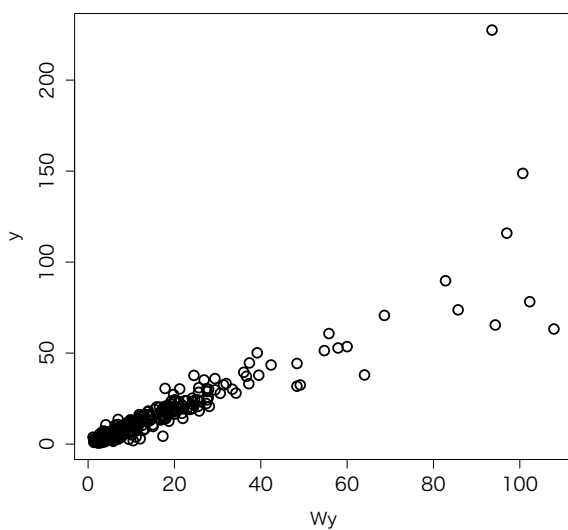


図 10.6 地価データ y と空間的従属性を考慮した地価データ W_y の散布図

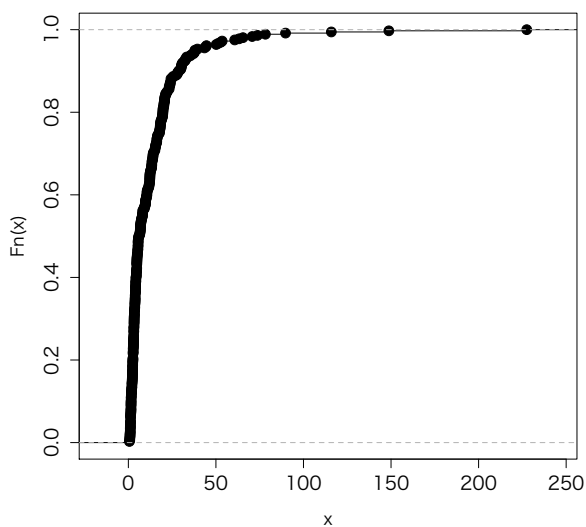


図 10.7 y の経験累積分布

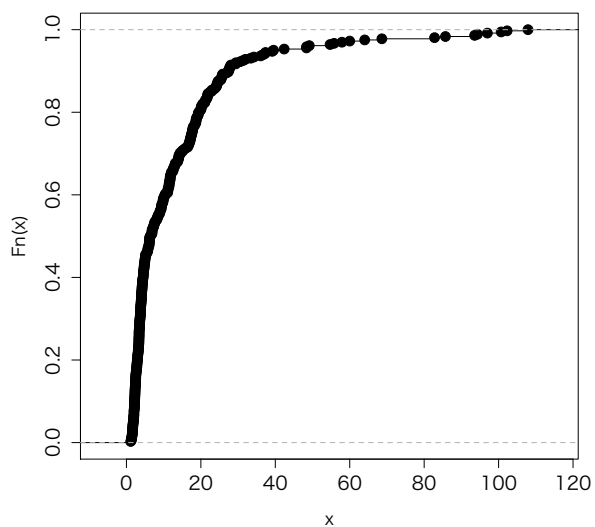


図 10.8 W_y の経験累積分布

表 10.5 空間的自己回帰モデルの推定結果

変数	最尤推定法		二段階最小二乗法	
	回帰係数	Z 値	回帰係数	t 値
定数項	1.71	4.50	1.27	2.36
夜間人口密度(POPD)	0.71	5.06	0.20	1.01
第三次産業従業人口密度(EMP3D)	1.60	15.49	1.26	9.18
ρ	0.44	8.41	0.67	8.37
		AIC=2504	残差分散=55.47	

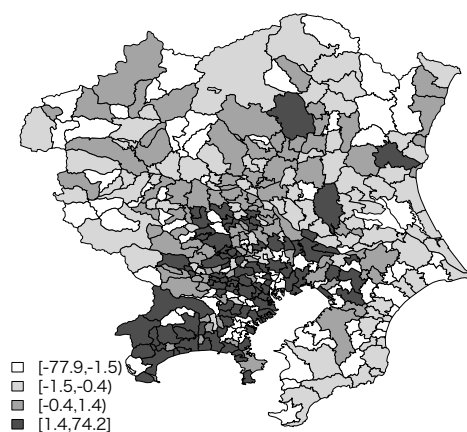


図 10.9 空間的自己回帰モデルの誤差項分布（最尤推定法）

表 10.6 空間的自己回帰モデルのベイズ推定結果

変数	平均	標準偏差	[2.5%, 97.5%]
定数項	1.41	0.54	[0.32, 2.43]
夜間人口密度(POPD)	0.37	0.18	[0.02, 0.71]
第三次産業従業人口密度(EMP3D)	1.34	0.13	[1.14, 1.46]
ρ	0.59	0.07	[0.46, 0.73]
σ^2	7.45	0.29	[6.93, 8.04]
τ	0.018	0.001	[0.015, 0.021]

DIC=2488.1

最尤法以外にも、二段階最小二乗法やベイズ法などによる推定方法も提案されている。二段階最小二乗法による推定結果(表 10.5)では、夜間人口密度(POPD)の回帰係数が 5%水準で統計的に有意となっていない。

空間的自己回帰モデルをベイズ推定するには、回帰係数の事前情報を正規分布、誤差項の分散 σ^2 の事前情報を逆ガンマ分布で与え、事後分布を MCMC 法で計算する。ここでは、事前情報を次式のように与えた場合の結果を示している。

$$\begin{aligned}y &= \rho W_i y_i + \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \\ \varepsilon_i &\sim N(\mu_i, \sigma^2) \\ \beta_0 &\sim N(0, 1 \times 10^{-6}) \\ \beta_1 &\sim N(0, 1 \times 10^{-6}) \\ \beta_2 &\sim N(0, 1 \times 10^{-6}) \\ \rho &\sim N(0, 1 \times 10^{-6}) \\ \sigma^2 &\sim IG(0.001, 0.001)\end{aligned}$$

マルコフ連鎖モンテカルロ法により空間的自己回帰モデルをベイズ推定した例を表 10.6 に示す。ここでは、MCMC の生成回数を 10,000 回、稼働検査期間を 1,000 回、チェーン数を 3 としている。

R 分析例

最尤法により空間的自己回帰モデルを推定してみよう。spdep パッケージの lagsarlm() 関数を用いて、空間的自己回帰モデルを推定できる。ここでは、市区町村代表点の座標値をもとにドロネー三角網図による隣接行列および空間重み付け行列を作成し、空間的自己回帰モデルの推定に用いている。既にパッケージとデータが読み込み済みであることを前提に、以下の手順でモデルを推定する。

lagsarlm() 関数の中で quiet=FALSE を指定すると、最尤推定法による収束過程が示される。

```
# 空間的自己回帰モデル
lph.lag <- lagsarlm(LPH~POPD+EMP3D, data=kanto,
nb2listw(lph.tri.nb, style="W"))
summary(lph.lag)
```

10.5.2 誤差項の空間的自己回帰モデル

誤差項の空間的自己回帰モデル（空間誤差モデル、spatial error model: SEM）は、次式のように誤差項に空間的自己相関を明示したモデルである。

$$\begin{aligned} y &= X\beta + u \\ u &= \lambda Wu + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I) \end{aligned}$$

ここで λWu は空間的従属性のある誤差項、 ε は空間的従属性のない誤差項、 λ は回帰係数を意味する。このモデルでは、誤差項の空間的従属性を、空間重み付け行列を用いて表現している。誤差項の分散共分散行列 Ω は、次式のようになる。

分散共分散行列 V は、次式のようになる。

$$V = (I - \lambda W)^{-1} \Sigma (I - \lambda W')^{-1}$$

誤差項の分散が均一 $\sigma_i^2 = \sigma^2$ であるとき、次のように置き換えることができる。

$$\Omega = \sigma^2 \left[(I - \lambda W)' (I - \lambda W) \right]^{-1}$$

地価データを用いた誤差項の空間的自己回帰モデルの最尤推定結果は、表 10.7 のようになる。誤差項の空間的従属性を示す回帰係数 λ は正であり、Z 値が統計的に有意であることが示された。また線形回帰モデルの AIC (2554.6) と比較して AIC が小さく、あてはまりがよいことがわかる。

表 10.4 の結果と比較すると、空間重み付け行列を用いて誤差項の空間的系列相関を考慮した同時自己回帰モデルと誤差項の空間的自己回帰モデルの推定結果は同じであることが理解できよう。

表 10.7 誤差項の空間的自己回帰モデルの推定結果

変数	最尤推定法		一般化モーメント法	
	回帰係数	Z 値	回帰係数	Z 値
定数項	3.00	4.26	2.81	4.34
夜間人口密度 (POPD)	1.58	12.73	1.63	13.90
第三次産業従業人口密度 (EMP3D)	2.18	25.31	2.21	26.56
λ	0.22	2.73 ¹⁾	0.14	3.99 ¹⁾
	AIC=2552.1		AIC=2552.6	

1) Z 値ではなく LR 検定統計量を示している

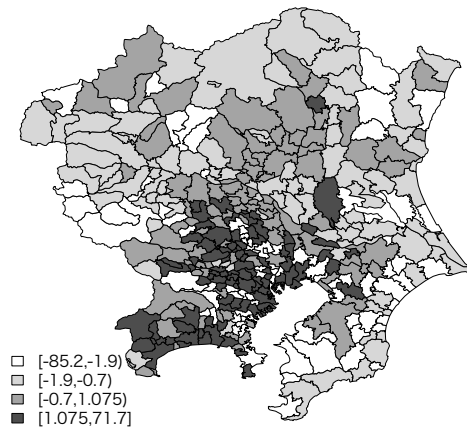


図10.10 誤差項の空間的自己回帰モデルの誤差分布（最尤推定法）

誤差項の空間的自己回帰モデルの誤差項分布を示したものが図 10.10 である。

誤差項の空間的自己回帰モデルは、最尤法の他に一般化モーメント法やベイズ法によりモデルを推定する方法が提案されている。

一般化モーメント法による推定結果を表 10.7 に示す。この方法は、 λ と σ^2 を同時に最適化する方法である。最尤推定法と比較して、最適化計算における数値探索局面があまり平坦でない（図 10.11）ため、最尤推定法の代替手段として用いられることもある。

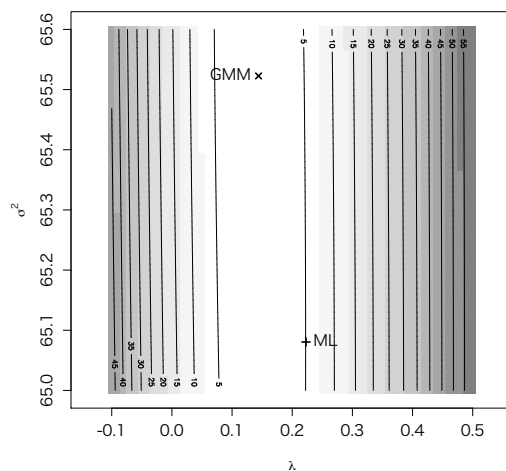


図 10.11 一般化モーメント法（GM）と最尤法（ML）の最適解

誤差項の空間的自己回帰モデルをベイズ推定するには、モデルを次式のように定式化する。ここでは、回帰係数の事前情報として正規分布、 λ の事前情報として一様分布 $U(0,1)$ 、誤差項の分散の事前情報として逆ガンマ分布を与えている。誤差項の空間的自己回帰モデルのベイズ推定結果を表 10.8 に示す。

$$\begin{aligned}
 y_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \\
 u_i &= \lambda W_i u_i + \varepsilon_i \\
 \varepsilon_i &\sim N(\mu_i, \sigma^2) \\
 \beta_0 &\sim N(0, 1 \times 10^{-6}) \\
 \beta_1 &\sim N(0, 1 \times 10^{-6}) \\
 \beta_2 &\sim N(0, 1 \times 10^{-6}) \\
 \lambda &\sim U(0, 1) \\
 \sigma^2 &\sim IG(0.001, 0.001)
 \end{aligned}$$

表 10.8 誤差項の空間的自己回帰モデルのベイズ推定結果

変数	平均	標準偏差	[2.5%, 97.5%]
定数項	-6.27	415.04	[-982.6, 964.4]
夜間人口密度(POPD)	-0.62	0.27	[-1.14, -0.07]
第三次産業従業人口密度(EMP3D)	1.39	0.13	[1.13, 1.64]
λ	0.99	0.013	[0.95, 1.00]
σ^2	7.51	0.28	[6.97, 8.09]
τ	0.018	0.001	[0.015, 0.021]

DIC=2491.4

R 分析例

誤差項の空間的自己回帰モデルは、`spdep` パッケージの `errorsarlm()` 関数を用いて推定することが出来る。

```

lph.err <- errorsarlm(LPH~POPD+EMP3D, data=kanto,
nb2listw(lph.tri.nb, style="W"))
summary(lph.err)

```

10.5.3 空間ダービンモデル

空間ダービンモデル (spatial Durbin model) は、説明変数と被説明変数の両方に空間的従属性を取り入れたモデルである。

$$y = \rho W y + X \beta + \rho W X \beta + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2 I)$$

地価データを用いて空間ダービンモデルを推定した例を表 10.9 に示す。この結果からは、夜間人口密度に対する回帰係数が負であり、Z 値が統計的に 5%水準で有意でないことがわかる。空間ダービンモデルの誤差分布を図 10.12 に示す。

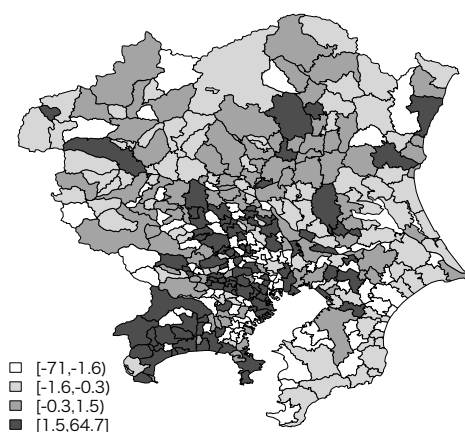


図 10.12 空間ダービンモデルの誤差項分布

表 10.9 空間ダービンモデルの最尤推定結果

変数	回帰係数	Z 値
定数項	1.35	2.43
夜間人口密度 (POPD)	-0.49	-1.83
第三次産業従業人口密度 (EMP3D)	1.44	11.36
ρ (Wy)	0.21	2.57
ρ (夜間人口密度)	1.88	5.75
ρ (第三次産業従業人口密度)	0.63	2.42

AIC=2475.4 (線形回帰モデルの AIC=2479.2)

空間ダービンモデルをベイズ推定するには、モデルを次式のように定式化し、事前情報を与える。

$$y_i = \rho W_i y_i + \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \rho W_i \beta_0 + \rho W_i \beta_1 x_{1i} + \rho W_i \beta_2 x_{2i} + \varepsilon_i$$

$$\varepsilon_i \sim N(\mu_i, \sigma^2)$$

$$\beta_0 \sim N(0, 1 \times 10^{-6})$$

$$\beta_1 \sim N(0, 1 \times 10^{-6})$$

$$\beta_2 \sim N(0, 1 \times 10^{-6})$$

$$\rho \sim N(0, 1 \times 10^{-6})$$

$$\sigma^2 \sim IG(0.001, 0.001)$$

空間ダービンモデルのベイズ推定結果を表 10.10 に示す。この結果から、夜間人口密度に対する回帰係数について、平均値が負となっているが、97.5%値は正となっている。

表 10.10 空間ダービンモデルのベイズ推定結果

変数	平均	標準偏差	[2.5%, 97.5%]
定数項	1.02	0.61	[-0.17, 2.23]
夜間人口密度(POPD)	-0.54	0.29	[-1.08, 0.03]
第三次産業従業人口密度(EMP3D)	1.41	0.13	[1.17, 1.50]
ρ (Wy)	0.39	0.14	[0.14, 0.65]
ρ (夜間人口密度)	1.59	0.37	[0.85, 2.31]
ρ (第三次産業従業人口密度)	0.22	0.36	[-0.48, 0.92]
σ^2	7.28	0.28	[6.76, 7.82]
τ	0.019	0.001	[0.016, 0.022]

DIC=2482.0

R 分析例

最尤法により空間ダービンモデルを推定するには、`spdep` パッケージの `lagsarlm()` 関数で引数 `type="mixed"` を指定する。

```
lph.durbin <- lagsarlm(LPH~POPD+EMP3D, data=kanto,
```

```
nb2listw(lph.tri.nb, style="W"), type="mixed")
summary(lph.durbin)
```

10.5.4 空間従属性の検定

観測データや誤差項に空間的従属性を取り入れるかどうかについては、空間的従属性を取り入れなかった場合、すなわち通常の線形回帰モデルと比較して、空間的従属性を考慮することの統計的有意性を判断する方法が提案されている。これを空間従属性に関するラグランジュ乗数検定という。

例えば、SARモデルの空間的従属性に関する回帰係数 ρ について、帰無仮説 H_0 と対立仮説 H_1 を次のようにおく。

帰無仮説 $H_0: \rho=0$

対立仮説 $H_1: \rho \neq 0$

このとき、次式で表されるラグランジュ乗数 LM について χ^2 検定を行い、有意水準 α に対して $LM > \chi^2(\alpha)$ であれば帰無仮説 H_0 を棄却し、空間的従属性を示す回帰係数 ρ を採用した方がよい、すなわちSARモデルを適用するのが望ましいと結論づける。

SARモデルの $\rho=0$ かどうかを検定した結果(LMlag)と、SEMモデルの $\lambda=0$ かどうかを検定した例(LMerr)を、表 10.11 に示す。この結果から、有意水準 5% で、SARモデルについては帰無仮説が棄却されるが、SEMモデルについては帰無仮説が棄却されないことがわかる。

表 10.11 空間的従属性の検定結果

モデル	ラグランジュ乗数 (LM)	p 値
SAR モデル (LMlag)	46.52	9.05×10^{-12}
SEM モデル (LMerr)	3.43	0.064

R 分析例

以下のようにして、SARモデルとSEMモデルの空間的従属性の有無に関する検定を行うことができる。

```
lm.LMtests(lph.lm, nb2listw(lph.tri.nb), test=c("LMerr", "LMlag"))
```

10.6 マルチレベルモデル

地域差を考慮して回帰モデルを推定する方法の一つに、マルチレベルモデルを適用することが考えられる。マルチレベルモデルは混合効果モデルとも呼ばれる。

夜間人口密度や第三次産業従業人口密度の分布が地価に与える影響は、特別区や政令指定都市、或いは県単位で異なると考えられる。そこで、ある地域グループ j （例えば、特別区、政令指定都市、特別区と政令指定都市を除く都県）ごとに回帰係数が異なるようなモデルを構築してみよう。

このとき、固定効果とランダム効果を考慮したモデルとして、以下のような組み合わせが考えられる。ここで、 y_{ij} は地域グループ j に属する市区町村 i の地価という意味である。

①切片と傾きが固定効果

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \varepsilon_{ij}$$

②切片がランダム効果で傾きが固定効果

$$y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \varepsilon_{ij}$$

③切片が固定効果で傾きがランダム効果

$$y_{ij} = \beta_0 + \beta_{1j} x_{1ij} + \beta_{2j} x_{2ij} + \varepsilon_{ij}$$

④切片と傾きがランダム効果

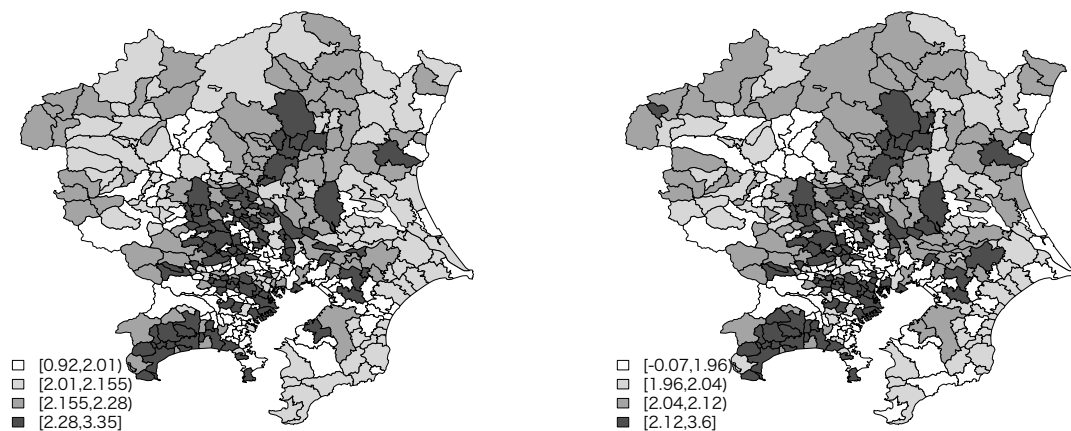
$$y_{ij} = \beta_{0j} + \beta_{1j} x_{1ij} + \beta_{2j} x_{2ij} + \varepsilon_{ij}$$

切片がランダム効果で傾きが固定効果のモデル②と、切片が固定効果で傾きがランダム効果のモデル③を推定した結果を、それぞれ表 10.12 に示す。

切片と傾きが共にランダム効果の場合のモデル④について、地域グループ j 単位ではなく、市区町村 i 単位で回帰係数を推定する場合、県・特別区・政令指定都市単位でモデルを推定する場合と比較して、より空間的な異質性を考慮したモデルとなる。即ち、以下のようなモデルとなる。

$$y_i = \beta_{0i} + \beta_{1i} x_{1ij} + \beta_{2i} x_{2ij} + \varepsilon_{ij}$$

このモデルは、しばしば最尤推定法では解が収束しないことがある。そこでベイズ推定した場合の、夜間人口密度及び第三次産業従業人口密度に対する回帰係数の分布を図 10.13 に示す。この結果から、それぞれの人口密度分布が地価に与える影響に地域差があることを理解できる。



(a) 夜間人口密度

(b) 第三次産業従業人口密度

図 10.13 市区町村単位でのバイズマルチレベルモデルの回帰係数推定結果

R 分析例

`lme4` パッケージの `lme()` 関数を用いて、モデル②とモデル③のマルチレベルモデルを推定する例を示す。新たに `lph.csv` データを用いる。

```
# lme4 パッケージを使用
library(lme4)
# lph データを読み込み
lph <- read.table("lph.csv", sep=",", header=T)
# モデル②（固定効果：傾き、ランダム効果：切片）を推定
lph.lme1 <- lme(LPH~POPD+EMP3D, random=~1|AREA, data=lph)
# 推定結果を表示
summary(lph.lme1)
# ランダム効果を表示
random.effects(lph.lme1)
# モデル③（固定効果：切片、ランダム効果：傾き）を推定
lph.lme2 <- lme(LPH~1, random=~0+POPD+EMP3D|AREA, data=lph)
# 推定結果を表示
```

```
summary(lph.lme2)
# ランダム効果を表示
random.effects(lph.lme2)
```

表 10.12 マルチレベルモデルの推定例

	変数	モデル②		モデル③	
		回帰係数	t 値	回帰係数	t 値
固定効果	定数項	4.60	3.43	-	-
	POPD	1.31	8.84	-	-
	EMD3D	2.12	25.38	3.60	5.66
ランダム効果	地域	定数項		POPD	EMD3D
	千葉県 ¹⁾	-2.23		1.51	1.68
	千葉市	-1.01		1.24	1.30
	群馬県	-2.78		0.027	0.036
	:	:		:	:
	東京都 23 区	7.88		1.76	2.25
	東京多摩地区	2.74		1.96	2.07
	横浜市	-1.31		1.52	0.78
AIC		2555.5		2578.7	

注 1) 千葉市を除く

10.7 地理的加重回帰モデル

空間的異質性と空間的従属性の両方を考慮した空間モデルとして、地理的加重回帰モデル (geographically weighted regression model: GWR) がある。このモデルは次式のように表され、地域 i ごとに異なる回帰係数 β_i と空間的従属性を示す空間重み付け関数 W_i とで構成される。

$$W_i y = W_i X \beta_i + \varepsilon_i$$

$$\varepsilon_i \sim N(0, \sigma^2 V_i)$$

ここで、 V_i は対角要素 $\{v_1, \dots, v_N\}$ で構成される対角行列である。

空間重み付け関数 W_i は、地域 i と他の地域との距離 d_i 及びバンド幅 θ を用いて、

以下のような距離低減関数が提案されている。

(1) 指数関数

$$W_i = \exp(-d_i/\theta)$$

(2) tri-cube 関数

$$W_i = \left(1 - (d_i/q_i)^3\right)^3$$

(3) ガウス関数

$$W_i = \exp\left(-\frac{1}{2}(d_i/\theta)^2\right)$$

(4) bi-square 関数

$$W_i = \exp\left(-(d_i/\theta)^2\right)$$

空間重み付け関数は以下の行列で構成される。

$$W_i = \begin{pmatrix} w_{i1} & 0 & \cdots & 0 \\ 0 & w_{i2} & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & w_{iN} \end{pmatrix}$$

またその要素 w_{ij} は、例えば指数関数を用いた場合には、次式のような基準化された値をとる。

$$w_{ij} = \frac{\exp(-d_{ij}/\theta)}{\sum_{j=1}^N \exp(-d_{ij}/\theta)}$$

指数関数やガウス関数で用いられるバンド幅 θ は、交差検証（クロスバリデーション）法を用いて以下のクロスバリデーションスコア CV を最小化することにより求められる。

$$CV = \sum_{i=1}^N [y_i - \hat{y}_{\neq i}(\theta)]$$

ここで $\hat{y}_{\neq i}(\theta)$ は i を除く観測データを用いた予測値である。

バンド幅 θ を変化させた場合のクロスバリデーションスコア CV は、[図 10.14](#) のような関係になる。

地価モデルに地理的加重回帰モデルを適用して得られた回帰係数の分布を [図](#)

10.15 に、誤差項の空間分布を 図 10.16 に、ローカルな R^2 値を 図 10.17 に、それぞれ示す。

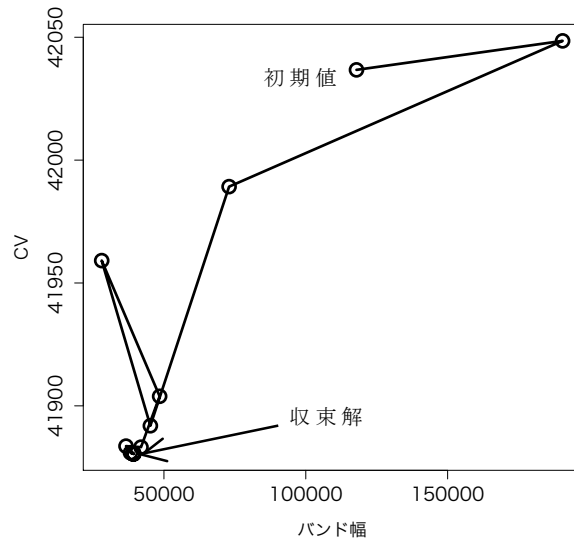
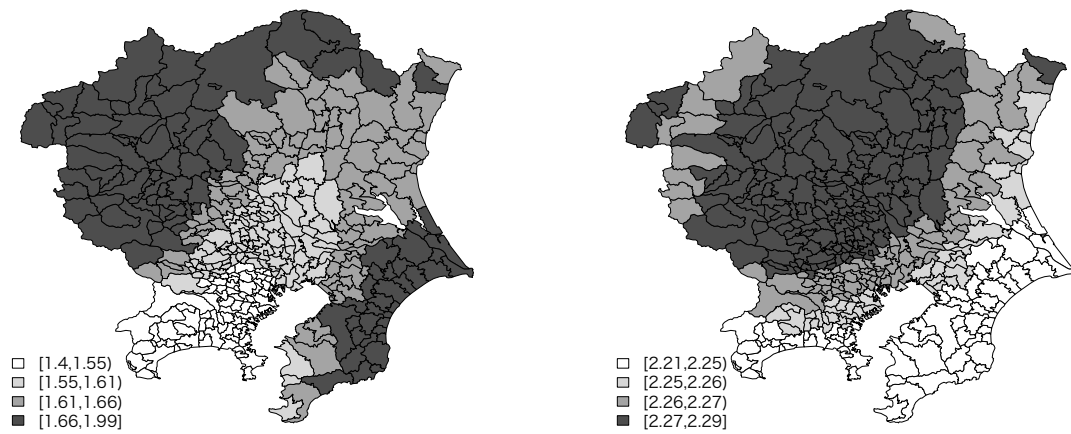


図 10.14 バンド幅 θ とクロスバリデーションスコア CV の軌跡



(a) 夜間人口密度

(b) 第三次産業従業人口密度

図 10.15 地理的加重回帰モデルのパラメータ推定結果

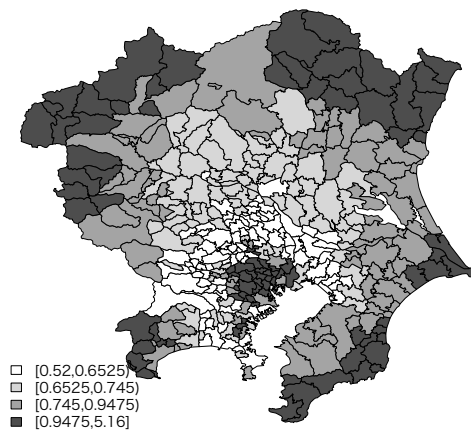


図 10.16 地理的加重回帰モデルの標準誤差分布

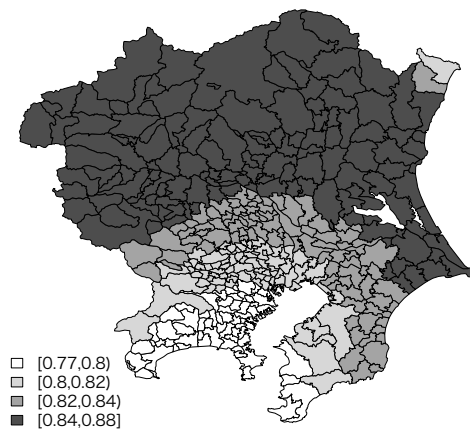


図 10.17 地理的加重回帰モデルのローカルな R^2 値

R 分析例

地理的加重回帰モデルを最尤推定してみよう。spgwr パッケージを用いて、まず `gwr.sel()` 関数によりバンド幅を計算する。得られたバンド幅を用いて、`gwr()` 関数により地理的加重回帰モデルを推定する。

```
# spgwr パッケージを使用
library(spgwr)
# バンド幅を計算
lph.bw <- gwr.sel(LPH~POPD+EMP3D, data=kanto, coords=coords)
# 地理的加重回帰モデルを推定
lph.gwr <- gwr(LPH~POPD+EMP3D, data=kanto, coords=coords,
bandwidth=lph.bw, hatmatrix=TRUE)
# モデル推定結果を表示
summary(lph.gwr$SDF)
```

参考文献

- [1] Anselin, L., R. Florax and S. Rey (eds.) (2004) *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer.

- [2] Arbia, G. (2006) *Spatial Econometrics*, Springer.
- [3] LeSage, J.P. and R.K. Pace (2009) *Introduction to Spatial Econometrics*, Chapman & Hall/CRC Press.
- [4] Fotheringham, A.S., C. Brunsdon and M. Charlton (2002) *Geographically Weighted Regression: the analysis of spatially varying relationships*, Wiley.
- [5] LeSage, J.P. (1999) *The Theory and Practice of Spatial Econometrics*, <http://www.spatial-econometrics.com/> .