

空間の統計学(8)： 地理的加重回帰モデル

慶應義塾大学総合政策学部准教授

古谷 知之 (Furutani Tomoyuki)

■兵庫県生まれ。2001年東京大学大学院工学系研究科博士課程修了。博士(工学)。東京大学大学院助手、慶應義塾大学環境情報学部専任講師を経て、07年4月より現職。専門分野：空間統計学、都市交通計画、観光政策。



1. はじめに

今回は、変数や誤差項の空間的自己相関を考慮した回帰モデルとして、誤差項の空間的自己回帰 (SEM) モデル、空間的自己回帰 (SAR) モデル、空間ダービンモデル、条件付き自己回帰 (CAR) モデルを紹介しました。いずれも、誤差項に分散均一な正規分布を仮定し、空間的に線形な現象を表現することを前提にした、固定効果のみを考慮した線形回帰モデルです。しかし空間データを扱ったモデリングを行う際には、地域差や空間的に非線形な現象についても考慮することが必要な場合があります。

そこで今回は、空間計量経済モデルを、より一般的に拡張したモデルを紹介します。一つは、空間的に非線形な現象を、線形回帰モデル同様に簡単に扱えるようにした、一般化線形回帰モデル及び一般化加法モデルと呼ばれるモデル群です。もう一つは、地域ごとにパラメータ推定を行うマルチレベルモデル、及び空間ウェイトを考慮する形で拡張した地

理的加重回帰モデルと呼ばれるモデル群です。マルチレベルモデルや地理的加重回帰モデルを用いれば、地域差を考慮した政策分析や予測が行えます。

演習には、前回同様、首都圏の市区町村別地価データ (住宅地標準地地価の平均価格) 及び夜間人口密度並びに第三次産業従業人口密度データを用いて、地価モデルの推定を例に挙げます。

今回、Rを使って、一般化線形モデルの推定にパッケージ `stat`、一般化加法モデルの推定に `mgcv` パッケージ、マルチレベルモデルの推定に `nlme` パッケージ、地理的加重回帰モデルの推定に `spgwr` パッケージを用います。マルチレベルモデルの推定には `lmer` パッケージを使うこともあります。

```
library(stats) # 一般化線形モデル
library(mgcv)  # 一般化加法モデル
library(nlme)  # マルチレベルモデル
library(spdep) # 空間データの操作
library(spgwr) # 地理的加重回帰モデル
```

データlph.csvには、市区町村の属性データとして、代表点座標、住宅地標準地地価(LPH) (万円/m²)、夜間人口密度(POPD) (千人/km²) 及び第三次産業従業人口密度(EMP3D) (千人/km²) が含まれています。図1～3に、住宅地地価、夜間人口密度、第三次産業従業人口密度の分布を、それぞれ図示します。また、今回はあらたに、市区町村可住地面積(km²)と都県名・政令指定都市名を使います。ただし、データの制約上、千葉市内の区については、区の面積を可住地面積として用いています。あらかじめ、空間隣接行列とドロネー三角網による空間重み付け行列を作成しておきましょう。

```
lph <- read.table("lph.csv",
sep=";", header=T)
summary(lph)
# 空間隣接行列の作成
coords <- as.matrix(
cbind(lph$Easting, lph$Northing))
nb <- tri2nb(coords)
# 空間重み付け行列の作成
nb.w <- nb2listw(nb, style="W")
```

また、通常の線形回帰モデルの結果は、以下の通りとなります (* 1)。

```
lph.lm <- lm(LPH~POPD+EMP3D,
data=lph)
summary(lph.lm)
```

2. 一般化線形回帰モデル

上述の線形回帰モデルでは、単位面積あたりの地価、夜間人口及び第三次産業従業人口を変数として扱いました。特に説明変数である夜間人口密度と第三次産業従業人口密度は、市区町村の可住地面積を人口で割ったデータを用いています。地区の面積規模による影響

図1 住宅地地価の分布

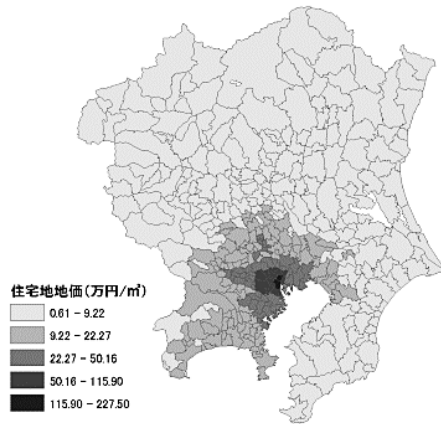


図2 夜間人口密度の分布

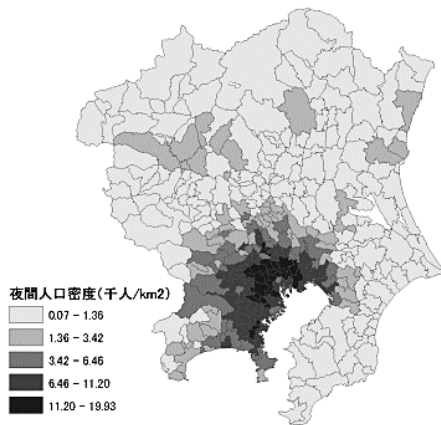
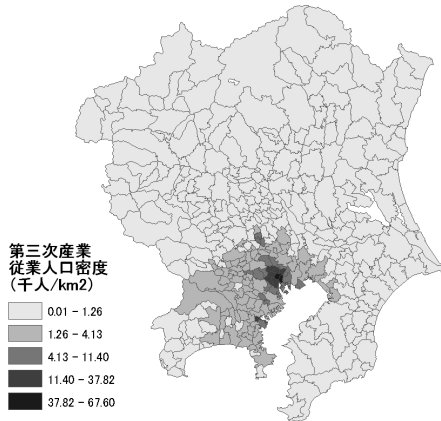


図3 第三次産業従業人口密度の分布



を考慮するため、オフセット項として面積に対する回帰係数を1に固定して回帰モデルを推定する方法が考えられます。

例えば、可住地面積の対数をとった非線形オフセット項を用いた一般化線形回帰モデル (generalized liner model) は、以下のように推定できます (* 2)。ここで、一般化線形回帰モデルは `glm()` 関数を用い、オフセット項は `offset()` 関数を用います。

```
lph.glm <- glm(LPH~POPD+EMP3D
+offset(log(S)), data=lph)
summary(lph.glm)
```

3. 一般化加法モデル

一般化加法モデル (generalized additive model) は、説明変数の変換を行う関数を組み込んだモデルです。

まず、平滑化関数 `s()` を用いて、緯度経度による平滑化関数を組み込んだモデルを推定します (* 3)。

```
lph.gam1 <- gam(LPH~POPD+EMP3D
+s(Easting, Northing), data=lph)
summary(lph.gam1)
```

次に、緯度経度による平滑化関数と可住地面積によるオフセット項を組み込んだモデルを推定します (* 4)。

```
lph.gam2 <- gam(LPH~POPD+EMP3D
+offset(log(S))+s(Easting, Northing),
data=lph)
summary(lph.gam2)
```

4. 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として、マルチレベルモデルと呼ばれる混合効果モデルを用いる方法があります。固定効果とランダム効果を考慮したモデルとして、

例えば、以下のような組み合わせのモデルが考えられます。

- ① 切片と説明変数が固定効果
- ② 切片がランダム効果で説明変数が固定効果
- ③ 切片が固定効果で説明変数がランダム効果
- ④ 切片と説明変数がランダム効果

例えば、切片がランダム効果で説明変数が固定効果の混合効果モデルは、以下のように推定できます。ここでは、地域区分として、政令指定都市及び政令指定都市以外の都県の12地域を用いることにします。nlmeパッケージでは `lme()` 関数を使ってマルチレベルモデルを推定することができます。

切片をランダム項、傾きを固定効果にする場合、以下の手順で推定します。また、`random.effects()` 関数を使って、ランダム項の推定結果を見ることができます (* 5)。

```
# 固定効果：傾き、ランダム効果：切片
lph.lme1 <- lme(LPH~POPD+EMP3D,
random=~1|AREA, data=lph)
summary(lph.lme1)
random.effects(lph.lme1)
```

傾きを固定効果、切片をランダム効果にする場合は、以下の手順で推定します (* 6)。

```
# 固定効果：切片、ランダム効果：傾き
lph.lme2 <- lme(LPH~1,
random=~0+POPD+EMP3D|AREA,
data=lph)
summary(lph.lme2)
random.effects(lph.lme2)
```

5. 地理的加重回帰モデル

空間近接性を取り込んだ混合効果モデルとして、地理的加重回帰モデルが知られていません。地理的加重回帰モデルは、小地域データなど地域差を表現すべきモデリングに有効と言われており、次式のように表されます。

$$W_i y_i = W_i X \beta_i + \varepsilon_i,$$

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

ここで、 i は地区、 y_i は被説明変数、 X は説明変数、 β_i は未知パラメータ、 V_i と σ はそれぞれ誤差項 ε_i に対するスケールパラメータとハイパーパラメータを意味します。地区数が n 個の場合、地区 i に対する空間重み付け行列 W_i は $n \times n$ 行列となります。

空間重み付け行列 W_i の要素 w_{ij} は地点 ij 間の距離 d_{ij} と距離減衰パラメータを意味するバンド幅 θ を用いて、例えば次式のように表されます。

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

θ は地理的加重回帰モデルの誤差項の平均平方根が最小となるように交差検定(cross-validation)により得られます。 w_{ij} はこのほかに、ガウス関数や二乗平方(bisquare)関数などが用いられます。

Rでは、`gwr.sel()`関数でバンド幅の交差検定スコアを算出し、その結果を基に`gwr()`関数を用いて地理的加重回帰モデルを推定します(*7)。

```
# バンド幅の計算
lph.bw <- gwr.sel(LPH~POPD+EMP3D,
data=lph, coords=coords)
# 地理的加重回帰モデルの推定
lph.gwr <- gwr(LPH~POPD+EMP3D,
data=lph, coords=coords,
bandwidth=lph.bw, hatmatrix=TRUE)
summary(lph.gwr$SDF)
```

地理的加重回帰モデルの夜間人口密度と第三次産業従業人口密度に対するパラメータ推定結果は、図4、図5のようになります。ま

図4 夜間人口密度に対するパラメータ分布

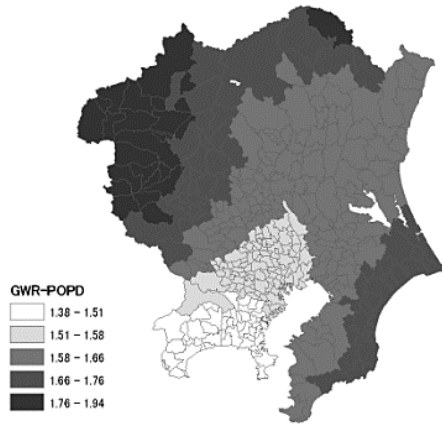


図5 第三次産業従業人口密度に対するパラメータ分布

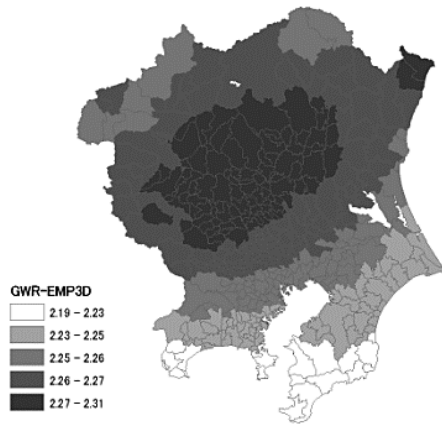
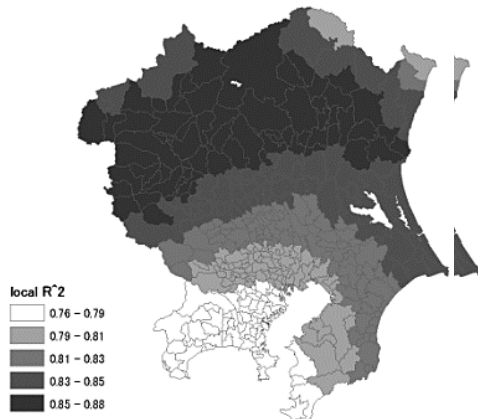


図6 ローカルなR²の分布



た、市区町村ごとのローカルな R^2 は図6のようになります。

* 参考文献

- [1] 金明哲 (2007) : Rによるデータサイエンス : 森北出版.
- [2] J. P. LeSage (2004) : Geographically Weighted Regression Models : Advances in Spatial Econometrics (L. Anselin, R.J.G.M. Florax and S.J. Rey eds.) : Springer.

* 1 線形回帰モデルlph.lmの推定結果

```
> summary(lph.lm)

Call:
lm(formula = LPH ~ POPD + EMP3D, data = lph)

Residuals:
    Min       1Q   Median       3Q      Max
-85.1101  -1.8732  -0.5808   1.2305  67.0602

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.55112    0.56647   4.504 9.05e-06 ***
POPD         1.68162    0.10611  15.848 < 2e-16 ***
EMP3D        2.24666    0.07841  28.654 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.188 on 359 degrees of freedom
Multiple R-squared:  0.8289, Adjusted R-squared:  0.8279
F-statistic: 869.6 on 2 and 359 DF, p-value: < 2.2e-16
```

* 2 一般化線形回帰モデルlph.glmの推定結果

```
> summary(lph.glm)

Call:
glm(formula = LPH ~ POPD + EMP3D + offset(log(S)), data = lph)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-84.9925  -2.1061  -0.7026   1.4599  67.4237

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.29070    0.56909  -2.268  0.0239 *
POPD         1.75745    0.10660  16.487 < 2e-16 ***
EMP3D        2.25733    0.07877  28.658 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 67.66322)

Null deviance: 144897 on 361 degrees of freedom
Residual deviance: 24291 on 359 degrees of freedom
AIC: 2558.0

Number of Fisher Scoring iterations: 2
```

* 3 一般化加法モデルlph.gam1の推定結果

```
> summary(lph.gam1)

Family: gaussian
Link function: identity

Formula:
LPH ~ POPD + EMP3D + s(Easting, Northing)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.56805    1.00229   8.548 4.55e-16 ***
POPD         0.21133    0.23183   0.912  0.363
EMP3D        2.01558    0.08435  23.897 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
            edf Ref.df    F p-value
s(Easting,Northing) 25.27  25.27  2.433 0.000199 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.849  Deviance explained = 86%
GCV score = 63.8  Scale est. = 58.818  n = 362
```

* 4 一般化加法モデルlph.gam2の推定結果

```
> summary(lph.gam2)
Family: gaussian
Link function: identity

Formula:
LPH ~ POPD + EMP3D + offset(log(S)) + s(Easting, Northing)

Parametric coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.93844    1.00134    4.932 1.29e-06 ***
POPD         0.23006    0.23165    0.993  0.321
EMP3D       2.02890    0.08417   24.105 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
      edf Ref.df  F  p-value
s(Easting,Northing) 25.39 25.39 2.682 3.40e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.85   Deviance explained = 86.6%
GCV score = 63.383   Scale est. = 58.412   n = 362
```

* 5 マルチレベルモデルlph.lme1の推定結果

```
> summary(lph.lme1)
Linear mixed-effects model fit by REML
Data: lph
      AIC      BIC    logLik
2555.529 2574.945 -1272.764

Random effects:
Formula: ~1 | AREA
(Intercept) Residual
StdDev:    3.409894 7.910093

Fixed effects: LPH ~ POPD + EMP3D
      Value Std.Error DF   t-value p-value
(Intercept) 4.600159 1.3422102 348  3.427301 7e-04
POPD        1.311962 0.1483348 348  8.844599 0e+00
EMP3D       2.121863 0.0835879 348 25.384815 0e+00

Correlation:
(Intr) POPD
POPD -0.543
EMP3D -0.121 -0.063

Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-10.54721777 -0.20096086 -0.05554292  0.12546537  8.45644210

Number of Observations: 362
Number of Groups: 12
> random.effects(lph.lme1)
(Intercept)
Chiba -2.2258620
Chiba-shi -1.0056714
Gunma -2.7770121
Ibaraki -2.4147557
Kanagawa 1.3627766
Kawasaki-shi 0.5765241
Saitama -0.4814151
Saitama-shi -0.4998201
Tochigi -1.8470578
Tokyo-ku 7.8808244
Tokyo-tama 2.7414956
Yokohama-shi -1.3100265
```

* 6 マルチレベルモデルlph.lme2の推定結果

```
> summary(lph.lme2)
Linear mixed-effects model fit by REML
Data: lph
      AIC      BIC    logLik
2578.708 2598.152 -1284.354

Random effects:
Formula: ~0 + POPD + EMP3D | AREA
Structure: General positive-definite, Log-Cholesky parametrization
      StdDev  Corr
POPD  1.496278 POPD
EMP3D  1.632696 0.902
Residual 7.930789

Fixed effects: LPH ~ 1
      Value Std.Error DF   t-value p-value
(Intercept) 3.604157 0.6368386 350  5.659451  0

Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-11.00181162 -0.26932644 -0.06655365  0.14637830  8.24555589

Number of Observations: 362
Number of Groups: 12
> random.effects(lph.lme2)
      POPD  EMP3D
Chiba  1.50629023 1.68993752
Chiba-shi 1.24286915 1.30469165
Gunma  0.02780028 0.03569207
Ibaraki 0.30044531 0.32388387
Kanagawa 1.79056653 1.85603303
Kawasaki-shi 1.49345210 1.64065941
Saitama 1.57533853 1.62873914
Saitama-shi 1.39561339 1.35610567
Tochigi 0.26048324 0.27283806
Tokyo-ku 1.76220965 2.25145795
Tokyo-tama 1.95960167 2.06945325
Yokohama-shi 1.52344912 0.78246809
```

* 7 地理的加重回帰モデルの推定結果

```
> summary(lph.gwr$SDF)
Object of class SpatialPointsDataFrame
Coordinates:
coord.x -120012.6 85196.34
coord.y 114091.4 115714.52
Is projected: NA
proj4string: [NA]
Number of points: 362
Data attributes:
      sum.w (Intercept) POPD EMP3D
Min.: 14.85 Min.: 1.114 Min.: 1.378 Min.: 2.191
1st Qu.: 71.48 1st Qu.: 2.265 1st Qu.: 1.556 1st Qu.: 2.250
Median: 114.71 Median: 3.110 Median: 1.610 Median: 2.264
Mean: 109.29 Mean: 3.158 Mean: 1.613 Mean: 2.260
3rd Qu.: 150.33 3rd Qu.: 4.006 3rd Qu.: 1.654 3rd Qu.: 2.274
Max.: 170.85 Max.: 5.659 Max.: 1.941 Max.: 2.302

(Intercept)_se POPD_se EMP3D_se gwr_e
Min.: 0.7196 Min.: 0.1163 Min.: 0.07789 Min.: -85.3098
1st Qu.: 0.7505 1st Qu.: 0.1186 1st Qu.: 0.07817 1st Qu.: -1.5152
Median: 0.8049 Median: 0.1241 Median: 0.07871 Median: -0.5827
Mean: 0.8794 Mean: 0.1508 Mean: 0.08415 Mean: -0.1439
3rd Qu.: 0.9506 3rd Qu.: 0.1393 3rd Qu.: 0.07992 3rd Qu.: 0.8215
Max.: 1.6788 Max.: 1.3407 Max.: 0.37839 Max.: 65.3154

pred localID (Intercept)_se_EDF
Min.: 1.353 Min.: 0.5168 Min.: 0.7689 Min.: 0.7222
1st Qu.: 3.742 1st Qu.: 0.6535 1st Qu.: 0.8023 1st Qu.: 0.7532
Median: 6.240 Median: 0.7391 Median: 0.8235 Median: 0.8078
Mean: 13.241 Mean: 0.8492 Mean: 0.8237 Mean: 0.8026
3rd Qu.: 18.461 3rd Qu.: 0.9491 3rd Qu.: 0.8467 3rd Qu.: 0.9540
Max.: 164.160 Max.: 5.1645 Max.: 0.8778 Max.: 1.6849

POPD_se_EDF EMP3D_se_EDF pred_se_EDF
Min.: 0.1167 Min.: 0.07817 Min.: 0.5187
1st Qu.: 0.1190 1st Qu.: 0.07845 1st Qu.: 0.6558
Median: 0.1245 Median: 0.07899 Median: 0.7418
Mean: 0.1514 Mean: 0.08446 Mean: 0.8523
3rd Qu.: 0.1398 3rd Qu.: 0.08021 3rd Qu.: 0.9526
Max.: 1.3455 Max.: 0.37976 Max.: 5.1832
```