

空間の統計学(7)： 空間的自己回帰モデル

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

今回から、面データや点データを使った空間計量経済モデルを紹介します。代表的な空間計量経済モデルとして、空間的自己回帰モデル(誤差項の空間的自己回帰モデル、説明変数や非説明変数に空間的自己回帰を取り入れたモデル、条件付き自己回帰モデル)や地理的加重回帰モデルなどが知られています。

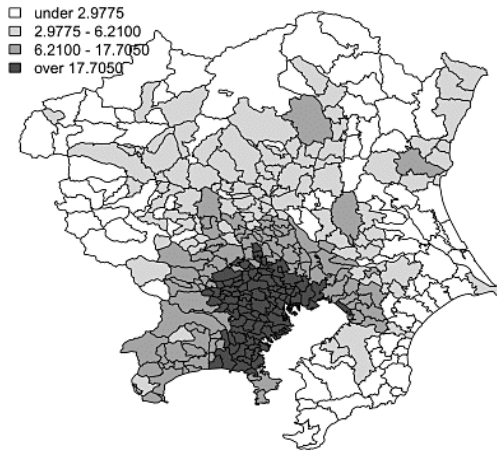
空間計量経済モデルは、空間疫学分野では標準化死亡比などの分析、都市経済学分野では地価など不動産市場の空間分析、生態学分野では生物の個体群・群集の分析、計量政治学分野では選挙行動の空間分析等に用いられます。

このうち今回は、空間的自己回帰モデルを紹介します。演習には、関東圏(一都五県)の地価データ(住宅地標準地地価の平均価格)と人口データ(夜間人口密度及び第三次産業従業人口密度)を用い[1]、人口データから地価データを説明する空間計量経済モデルを作成します。

Rでは、パッケージspdepを使って、空間的自己回帰モデルを推定することができます。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に地価分布の主題図を示します。主題図を作成するために、パッケージmapprojとclassIntを使います。

```
library(spdep)
library(mapproj)
library(classInt)
lph <- read.table("lph.csv", sep=";", header=T)
summary(lph)
kanto <- readShapePoly("kanto_area.shp",
IDvar="JCODE")
# 地図表示
pal1 <- gray.colors(n=4,start=1,end=0.3)
q_kanto <- classIntervals(kanto$LPH, n=4,
style="quantile")
q_kanto_Col <- findColours(q_kanto,pal1)
plot(kanto,col=q_kanto_Col)
legend("topleft",fill=attr(q_kanto_Col,"palette"),
legend=names(attr(q_kanto_Col,"table")),
bty="n")
```

図1 地価データの分布



2. 空間的自己相関

(1) 変数の空間的自己相関

線形回帰モデルに空間依存性を取り入れる前に、各変数の空間的自己相関性を示すことにしましょう。ここでは、ドロネー三角網による空間隣接行列を作成し、Moran's Iを計算します。

```
coords <- matrix(0, nrow=nrow(lph), ncol=2)
coords[,1] <- lph$Easting
coords[,2] <- lph$Northing
lph.tri.nb <- tri2nb(coords)
moran.test(lph$LPH,nb2listw(lph.tri.nb,style="W"))
moran.test(lph$POPD,nb2listw(lph.tri.nb,style="W"))
moran.test(lph$EMP3D,nb2listw(lph.tri.nb,style="W"))
```

この結果、住宅地平均地価、夜間人口密度、第三次産業従業人口密度のMoran's I値はそれぞれ0.76、0.86、0.63となり、空間的自己相関性が強いことがわかります。

次に、線形回帰モデルを推定し、誤差項の空間的自己相関を計算しましょう。

```
lph.lm <- lm(LPH~POPD+EMP3D,data=lph)
summary(lph.lm)
lm.morantest(lph.lm, nb2listw(lph.tri.nb, style="W"))
```

線形回帰モデルの推定結果

```
> 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
```

この結果、Moran's I値は0.057となり、誤差項の空間的自己相関は弱い（ない）ことがわかります。

(2) 空間依存性の検定

線形回帰モデルに空間依存性を取り入れるには、誤差項・説明変数・非説明変数に空間依存性を取り入れる方法が考えられます。線形回帰モデルに空間依存性を取り入れることが必要であるかどうかを判断するには、誤差項や変数の空間的自己相関の強弱を示すだけでは十分ではなく、空間依存性を考えない場合のモデルの統計的有意性を示すことが必要です。lm.LMtest()関数を用いれば、誤差項

や変数の空間依存性に関するラグランジュ乗数検定を行うことができます。

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

ここで、LMerrは誤差項の空間依存性に関する検定、LMlagは空間依存変数がない場合の検定、RLMerrは空間依存変数がない場合の誤差項の空間依存性に関する検定、RLMlagは空間依存変数がある場合の誤差項の空間依存性に関する検定の結果です。この結果から、空間依存変数があるモデルがより望ましいことがわかります。

lm.LMtest()の結果

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

Lagrange multiplier diagnostics for spatial
dependence

data:
model: lm(formula = LPH ~ POPD + EMP3D, data = lph)
weights: nb2listw(lph.tri.nb)

LMerr = 3.4311, df = 1, p-value = 0.06398

Lagrange multiplier diagnostics for spatial
dependence

data:
model: lm(formula = LPH ~ POPD + EMP3D, data = lph)
weights: nb2listw(lph.tri.nb)

LMlag = 46.524, df = 1, p-value = 9.05e-12

Lagrange multiplier diagnostics for spatial
dependence

data:
model: lm(formula = LPH ~ POPD + EMP3D, data = lph)
weights: nb2listw(lph.tri.nb)

RLMerr = 15.7534, df = 1, p-value = 7.216e-05
```

```
Lagrange multiplier diagnostics for spatial
dependence

data:
model: lm(formula = LPH ~ POPD + EMP3D, data = lph)
weights: nb2listw(lph.tri.nb)

RLMlag = 58.8462, df = 1, p-value = 1.710e-14

Lagrange multiplier diagnostics for spatial
dependence

data:
model: lm(formula = LPH ~ POPD + EMP3D, data = lph)
weights: nb2listw(lph.tri.nb)

SARMA = 62.2774, df = 2, p-value = 2.998e-14
```

3. 空間的自己回帰モデル

(1) 誤差項の空間的自己回帰 (SEM) モデル

誤差項の空間的自己回帰モデル (空間同時自己回帰誤差モデル) は、次式のように誤差項に空間的自己回帰を明示したモデルです。ここで、 y は非説明変数、 X は説明変数、 β と λ は未知パラメータ、 ϵ は空間依存しない誤差項、 λWu は空間依存する誤差項、 W は空間重み付け行列、 Σ は誤差項の分散行列を意味します。

$$y = X\beta + u$$

$$u = \lambda Wu + \epsilon$$

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

Rでは、`errorsarlm()`関数を用いて誤差項の空間的自己回帰モデルを最尤推定できます。

```
lph.err <- errorsarlm(LPH~POPD+EMP3D,
data=lph,
nb2listw(lph.tri.nb, style="W"),
method="eigen", quiet=FALSE)
summary(lph.err)
```

誤差項の空間的自己回帰モデルの推定結果

```
> summary(lph.err)

Call:
errorsarlm(formula = LPH ~ POPD + EMP3D, data = lph,
listw = nb2listw(lph.tri.nb,
style = "W"), method = "eigen", quiet = FALSE)

Residuals:
      Min       1Q   Median       3Q      Max
-85.20264 -1.85360 -0.67849  1.06294  71.74931

Type: error
Coefficients: (asymptotic standard errors)
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.003716   0.705467  4.2578 2.065e-05
POPD         1.583495   0.124412 12.7278 < 2.2e-16
EMP3D        2.184385   0.086304 25.3105 < 2.2e-16

Lambda: 0.22282 LR test value: 4.5444 p-value:
0.033026
Asymptotic standard error: 0.081399 z-value: 2.7374
p-value: 0.0061923
Wald statistic: 7.4935 p-value: 0.0061923

Log likelihood: -1271.043 for error model
ML residual variance (sigma squared): 65.081, (sigma:
8.0672)
Number of observations: 362
Number of parameters estimated: 5
AIC: 2552.1, (AIC for lm: 2554.6)
```

空間的自己回帰モデルの推定結果 (最尤推定法)

```
> summary(lph.lag)

Call:
lagsarlm(formula = LPH ~ POPD + EMP3D, data = lph,
listw = nb2listw(lph.tri.nb,
style = "W"), method = "eigen", quiet = FALSE)

Residuals:
      Min       1Q   Median       3Q      Max
-77.9456  -1.4924  -0.4296   1.4047  74.2277

Type: lag
Coefficients: (asymptotic standard errors)
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.70589    0.53639   3.1803 0.001471
POPD         0.70913    0.14015   5.0597 4.199e-07
EMP3D        1.59916    0.10322  15.4929 < 2.2e-16

Rho: 0.4375 LR test value: 52.582 p-value: 4.1267e-13
Asymptotic standard error: 0.052036 z-value: 8.4076
p-value: < 2.22e-16
Wald statistic: 70.687 p-value: < 2.22e-16

Log likelihood: -1247.024 for lag model
ML residual variance (sigma squared): 55.413, (sigma:
7.444)
Number of observations: 362
Number of parameters estimated: 5
AIC: 2504, (AIC for lm: 2554.6)
LM test for residual autocorrelation
test value: 8.4034 p-value: 0.0037452
```

(2) 空間的自己回帰 (SAR) モデル

空間的自己回帰モデル (空間同時自己回帰ラグモデル) は、次式のように非説明変数の空間ラグ依存を明示したモデルです。

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

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

Rでは、`lagsarlm()`関数を用いて空間的自己回帰モデルを最尤推定できます。

```
lph.lag <- lagsarlm(LPH~POPD+EMP3D,
data=lph,
nb2listw(lph.tri.nb, style="W"),
method="eigen", quiet=FALSE)
summary(lph.lag)
```

空間ラグは、`lag.listw()`関数を用いて計算することができます。図2にオリジナルな地価データと空間ラグを取り入れた地価データの関係を表す散布図を示します。また図3と図4は、それぞれオリジナルな地価と空間依存を考慮した地価の累積分布図です。

```
LPH.lag <- lag.listw(nb2listw(lph.tri.nb,
style="W"), lph$LPH)
plot(LPH.lag, lph$LPH)
plot(ecdf(lph$LPH))
plot(ecdf(LPH.lag))
```

図2 オリジナルな地価データと空間ラグ地価データとの相関関係

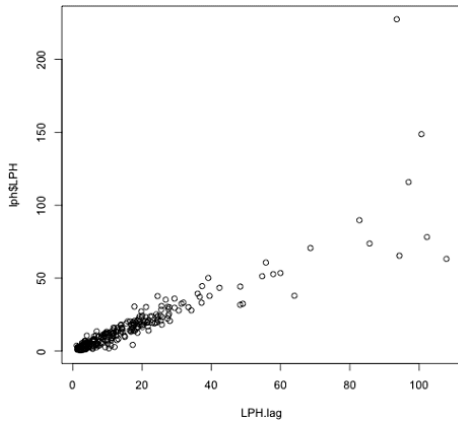


図3 地価データの経験累積分布

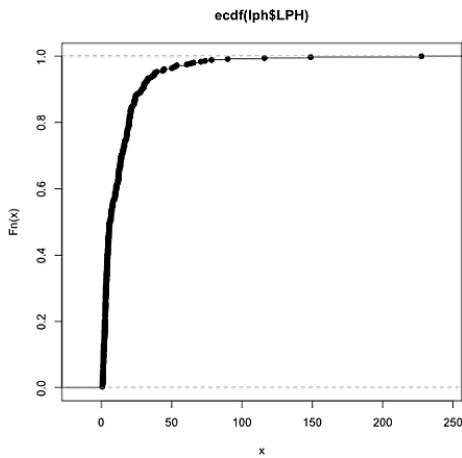
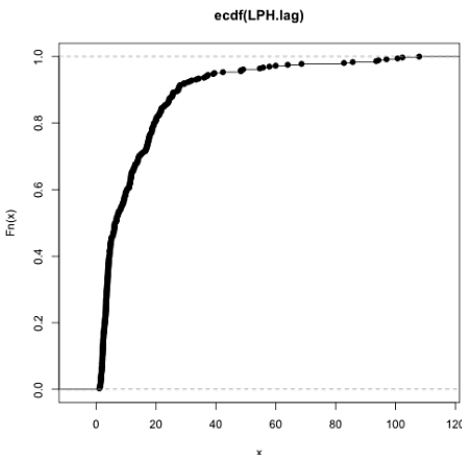


図4 空間ラグ地価データの経験累積分布



`lm()`関数を使って、空間的自己回帰モデルを最小二乗推定することができます。最尤推定の結果と比較すると、最小二乗推定した場合は夜間人口密度 (POPD) に対するパラメータが5%水準で有意でないことがわかります。

```
lph.lag2 <- lm(lph$LPH~LPH.lag+lph$POPD+lph$EMP3D)
summary(lph.lag2)
```

空間的自己回帰モデルの推定結果 (最小二乗法)

```
> summary(lph.lag2)

Call:
lm(formula = lph$LPH ~ LPH.lag + lph$POPD + lph$EMP3D)

Residuals:
    Min       1Q   Median       3Q      Max
-75.4972  -1.4950  -0.4260   1.5274   76.6772

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.41703    0.53014   2.673  0.00786 **
LPH.lag      0.58701    0.06664   8.808 < 2e-16 ***
lph$POPD     0.37679    0.17671   2.132  0.03366 *
lph$EMP3D    1.37788    0.12164  11.328 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.433 on 358 degrees of freedom
Multiple R-squared:  0.8594, Adjusted R-squared:  0.8582
F-statistic: 729.3 on 3 and 358 DF, p-value: < 2.2e-16
```

(3) 空間ダービンモデル

空間ダービン (Spatial Durbin) モデルは、説明変数と非説明変数の両方に空間ラグを取り入れたモデルです。

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

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

`lagsarlm()`関数に`type="mixed"`というオプションを追加することにより、Spatial Durbinモデルを推定できます。

```
lph.durbin <- lagsarlm(LPH~POPD+EMP3D,
data=lph,nb2listw(lph.tri.nb, style="W"),
type="mixed", method="eigen", quiet=FALSE)
summary(lph.durbin)
```

空間ダービンモデルの推定結果

```
> summary(lph.durbin)

Call:
lagsarlm(formula = LPH ~ POPD + EMP3D, data = lph,
listw = nb2listw(lph.tri.nb,
style = "W"), type = "mixed", method = "eigen",
quiet = FALSE)

Residuals:
    Min       1Q   Median       3Q      Max
-70.9760  -1.5784  -0.2734   1.5379   64.7118

Type: mixed
Coefficients: (asymptotic standard errors)
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.35401    0.55709   2.4305  0.01508
POPD         -0.49280    0.26789  -1.8396  0.06583
EMP3D        1.44224    0.12693  11.3627 < 2.2e-16
lag.POPD     1.88245    0.32732   5.7511  8.865e-09
lag.EMP3D    0.63268    0.26172   2.4174  0.01563

Rho: 0.20548 LR test value: 5.7358 p-value: 0.016623
Asymptotic standard error: 0.079946 z-value: 2.5703
p-value: 0.010162
Wald statistic: 6.6063 p-value: 0.010162

Log likelihood: -1230.722 for mixed model
ML residual variance (sigma squared): 52.155, (sigma:
7.2218)
Number of observations: 362
Number of parameters estimated: 7
AIC: 2475.4, (AIC for lm: 2479.2)
LM test for residual autocorrelation
test value: 1.1969 p-value: 0.27394
```

また、誤差項の空間的自己回帰モデル、空間的自己回帰モデル、Spatial Durbinモデルについて、これらの同時自己回帰モデルは **predict()**関数を用いて予測値を返すことができます。

```
predict(lph.err)
predict(lph.lag)
predict(lph.durbin)
```

(4) 空間的自己相関を考慮した条件付き自己回帰 (CAR) モデル

空間的自己相関を考慮した条件付き自己回帰モデルは、誤差項の分散がSARモデルとは、次のように異なります。

$$\text{SARモデル: } \Sigma = (I - \rho W)^{-1} D \sigma^2$$

$$\text{CARモデル: } \Sigma = (I - \rho W)^T D (I - \rho W)^{-1} \sigma^2$$

spautolm()関数を用いて、CARモデルを推定できます。関数の中で、family="CAR"と指定します。

```
lph.car <- spautolm(LPH~POPD+EMP3D,
data=lph, nb2listw(lph.tri.nb, style="W"),
family="CAR", method="full")
summary(lph.car)
```

条件付き自己回帰モデルの推定結果

```
> summary(lph.car)

Call:
spautolm(formula = LPH ~ POPD + EMP3D, data = lph,
listw = nb2listw(lph.tri.nb,
style = "W"), family = "CAR", method = "full")

Residuals:
    Min       1Q   Median       3Q      Max
-88.03571  -1.57639  -0.60198   1.08564  73.70468

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.22980    0.72637   4.4465  8.728e-06
POPD         1.54365    0.12666  12.1873 < 2.2e-16
EMP3D        2.15405    0.08679  24.8191 < 2.2e-16

Lambda: 0.43973 LR test value: 4.5286 p-value:
0.033332

Log likelihood: -1271.051
ML residual variance (sigma squared): 64.446, (sigma:
8.0278)
Number of observations: 362
Number of parameters estimated: 5
AIC: 2552.1
```

*参考文献

- [1] 総務省統計局：統計でみる市町村のすがた2009 (<http://www.stat.go.jp/data/ssds/5b.htm>) .