第 5 章　空間的自己相関

この章では、空間データの統計分析を行う上で基本となる、空間オブジェクト間の隣接関係を定義する**空間隣接行列**と、それに基づく**空間重み付け行列**について解説する。その上で、空間オブジェクトの属性データについての**空間的自己相関**の有無を検出する指標を紹介する。

自己相関（Autocorrelation）は、時系列解析においてよく知られる概念である。時系列自己相関のアナロジーから、空間データの系列相関を扱う際にも用いられる。空間的自己相関とは、おおまかにいえば、空間オブジェクトの属性データが、互いに近い地域・地点同士で似たような値を示す傾向があるか、それともランダムに分布する傾向があるかを示す指標である。点オブジェクトや面オブジェクトに対して、このような分析を適用することが多い。

空間的自己相関を計算するためには、空間隣接行列と空間重み付け行列を用いる。空間隣接行列は、空間オブジェクト間の隣接関係を表す。空間重み付け行列は、空間オブジェクト間の隣接関係や距離などのインピーダンス指標を元に、「近さ」を表現するための指標である。空間的自己相関を示す指標には、Join count 統計量、Moran's I、Geary's C、G 統計量、Local Moran's I などがある。

5.1 空間隣接行列

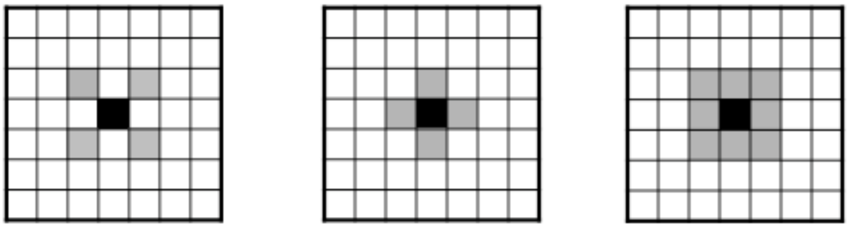
空間オブジェクトの属性データが空間的な系列相関を持つかどうかを検出するために、また空間的な自己相関を考慮した空間計量経済モデルを構築する上で、空間オブジェクト間の隣接関係を定義することは重要である。空間統計学では、空間隣接行列を用いて隣接関係を示すことができる。

空間的な隣接関係の表現方法は、空間オブジェクトの種類（グリッドデータ、点オブジェクト、面オブジェクト）によって異なる。ここではまず、グリッドデータでの隣接関係を表現する方法を解説する。次に、点オブジェクトや面オブジェクトでの隣接関係を表現する際に用いられる方法を紹介する。具体的には、ドロネー三角網を用いる方法、隣接地点（地区）数を指定する方法、一定半径以内に含まれる地点（地区）について隣接関係を定義する方法、について説明する。

5.1.1 グリッドデータの隣接行列

まず、グリッドデータの隣接関係の表現方法をみてみよう。グリッドデータの

場合、例えばチェスのアナロジーで、図 5.1 のように、黒塗りされたグリッドに対して灰色のグリッドが隣接しているというように隣接関係を決める方法がある。



(a) ルーク型　(b) エッジ型　(c) クイーン型

図 5.1 グリッドデータの隣接関係の定義方法

図 5.1 のうち、(a)のように上下左右に接するグリッドのみを定義する方法をルーク型隣接関係、(b)のように角に接するグリッドのみ隣接関係を定義する場合をエッジ型（ビショップ型）隣接関係、(c)のように上下左右と角が接するグリッドについて隣接関係を定義する場合をクイーン型隣接関係、等と呼ぶことがある。

次に、単純なグリッドデータとして、図 5.2 のような３×３の格子状の 9 区画を例に挙げて、ルーク型の隣接関係を空間隣接行列として表現してみよう。このとき、当該メッシュに対して上下左右の辺が接しているメッシュについては「接している（＝１）」、そうでない場合は「接していない（＝０）」と定義することにすると、空間隣接行列Cは次のように表すことができる。

x_1	x_2	x_3
x_4	x_5	x_6
x_7	x_8	x_9

図 5.2 ３×３の格子状区画

$$
C = \begin{pmatrix}
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0
\end{pmatrix}
$$

5.1.2 ドロネー三角網

点オブジェクトの場合、「最も近い地点」同士を結んで隣接関係を定義することができる、面オブジェクトの場合でも、ポリゴンの代表地点を定義することにより、点オブジェクトと同じように隣接関係を定義できる。

地点同士をつなぎ、隣接関係を示す方法の一つに、ドロネー三角網という方法がある。当該点から近い点同士を結び、三角形を形成する方法である。横浜市の区境界データの代表点座標を用いてドロネー三角網図を作成したものが図 5.3 である。ドロネー三角網を使った場合は、区境界が隣接しているかどうかにかかわらず、隣接関係が定義される。また、地点数が密に分布する地域では、より密にネットワークが形成されることがわかる。

R 分析例

```
# spdep パッケージを使用
library(spdep)
# 地図データの読み込み
yoko<- readShapePoly("yoko.shp", IDvar="JCODE")
# 座標テーブルの作成
yoko_coords <- coordinates(yoko)
# ドロネー三角網図の作成
yoko.tri.nb <- tri2nb(yoko_coords)
# ポリゴンデータの図示
plot(yoko, border="white", col="grey")
```

```
# ドロネー三角網図の図示
plot(yoko.tri.nb, yoko_coords, add=T)
```

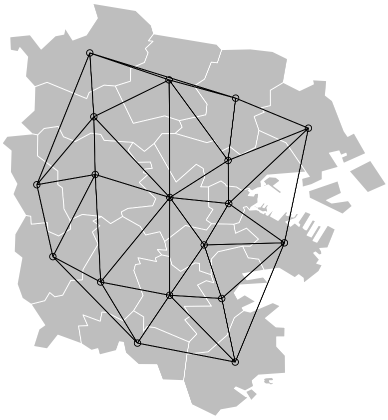


図 5.3 ドロネー三角網図

5.1.3 最近隣 k 地点を隣接関係と定義する方法

地点間の距離を元に、任意の空間オブジェクトから、最も近い k 地点のみを抽出し、それらを隣接地点と定義する方法である。横浜市区の代表点座標を用いて、代表点から最近隣 4 地点（$k=4$）について隣接関係を定義したものを図 5.4 に示す。また、$k=3$ の場合との違いを図 5.5 に示す。

R による分析例

```
# 再近隣 k=4 地点で隣接関係を定義
yoko.knn4 <- knearneigh(yoko_coords, k=4)
yoko.knn4.nb <- knn2nb(yoko.knn4, row.names=rownames(yoko$JCODE))
# 結果を図示
plot(yoko, border="white", col="grey")
plot(yoko.knn4.nb, yoko_coords, add=TRUE)
# 再近隣 k=3 地点で隣接関係を定義
kanto.knn3 <- knearneigh(coords, k=3)
```

```
yoko.knn3.nb <- knn2nb(yoko.knn3, row.names=rownames(yoko$JCODE))
# k=3 と K=4 の違い
diffs <- diffnb(yoko.knn3.nb, yoko.knn4.nb)
# k=3 と K=4 の違いを赤破線で図示
plot(yoko, border="white", col="grey")
plot(yoko.knn3.nb, yoko_coords, add=TRUE)
plot(diffs, yoko_coords, col="red", lty=2, lwd=2, add=TRUE)
```

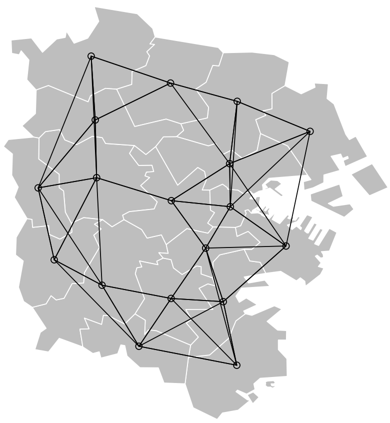


図 5.4　$k=4$のときの隣接関係

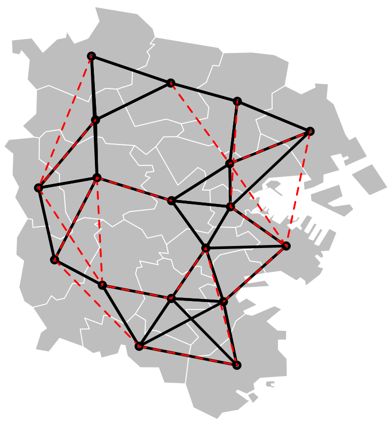


図 5.5　$k=3$と $k=4$の隣接関係の違い

5.1.4 距離により隣接関係を定義する方法

ある一定のユークリッド距離 rを半径とする円の内部に含まれる地点についてのみ隣接関係を定義する方法である。半径 r以内に隣接地点が一つも含まれない地点は、隣接関係が定義されないことになる。またその応用として、上限値 r_1を半径とする円のうち、下限値 r_2を半径とする円をドーナツ状にくりぬいた領域に含まれる地点についてのみ隣接関係を定義する方法もある。関東地方の市区町村中心点座標をもとに、中心点から半径 $r=0.5$として隣接関係を定義したものを図 5.6 に示す。この図からはやや判読しにくいが、自治体が密に分布する地域では隣接関係が密に定義されている一方で、自治体の面積規模が大きい郊外部では隣接関係が定義されない自治体もあることがわかる。

R 分析例

```
# 距離 r=0.06 の範囲内で隣接関係を定義
yoko.r.nb <- dnearneigh(yoko_coords, 0, 0.06)
# 結果を図示
plot(yoko, border="white", col="grey")
plot(yoko.r.nb, yoko_coords, add=TRUE)
```

5.1.6 面オブジェクトの隣接関係

面オブジェクトの境界同士が隣接しているかどうかで隣接関係を定義する方法である。関東地方の市区町村境界データをもとに隣接関係を定義したものを、図 5.7 に示す。

R 分析例

```
# 面オブジェクトの隣接行列を定義
yoko.poly.nb <- poly2nb(yoko)
# 結果を図示
plot(yoko, border="white", col="grey")
plot(yoko.poly.nb, yoko_coords, add=TRUE)
```

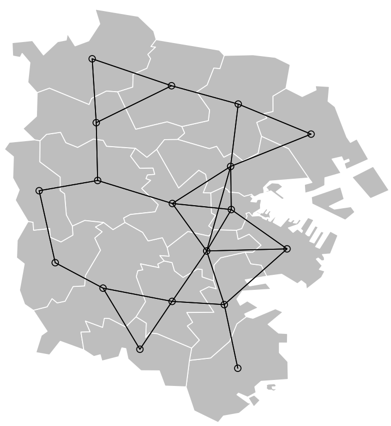


図 5.6 距離による隣接関係の定義（半径 $r = 0.06$）

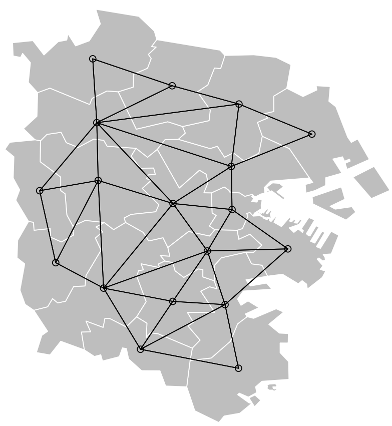


図 5.7 面データの境界による隣接関係の定義

5.2 空間重み付け行列

隣接関係が定義されれば、隣接関係を重み付け化して地点（地区）間の近接性を表現できる。これを空間重み付け行列という。

空間重み付け行列には、(１) 隣接行列をそのまま用いる方法 (W_B)、(２) 隣接行列の行和で標準化する方法 (W_W)、(３) 隣接行列の全要素の和で標準化する方法 (W_C)、(４) 距離行列を用いて標準化する方法 (W_S)、などがある。

ふたたび、図 5.2 のようなグリッド上の 9 区画を対象に、空間重み付け行列を作ることを考えよう。要素 $c_{ij}(i, j = 1,\ldots,9)$からなる空間隣接行列 Cが与えられたとき、隣接行列の行和で標準化する W_Wの要素 w_{ij}は次式のように表される。

$$w_{ij} = \frac{c_{ij}}{\sum_{i=1}^{9} c_{ij}}$$

このとき、空間重み付け行列 W_Wは、以下のようになる。

$$W_W = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 \\ 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \end{pmatrix}$$

5.3 空間的自己相関分析

空間的自己相関を示す指標のうち、Moran's I、Geary's C、Join 統計量は、対象地域全域でのグローバルな自己相関を示す指標として知られる。このうち Geary's C は Moran's I よりも局所的な自己相関を示す。G 統計量や Local Moran's I は、局地的な空間的な自己相関を示す指標として、ホットスポットを検出する際などに用いられる。

5.3.1 Moran's I

N地区（地点）からなる対象地域において、地区 iの属性 x_iをとする。対象地域全体における属性の平均値を $\bar{x}$、地点 ij間の空間重み付け行列の要素を w_{ij}とすると、

Moran's I は次式のように表される。

$$Moran's\ \ I = \frac{N}{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}} \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

この式に示されているように、Moran's I は相関係数に空間重み付け行列の要素を考慮した指標であるといえる。Moran's I は、−1 から 1 の間の数値をとります。1 に近い値の時、互いに近い空間オブジェクトの属性 x_iが類似するような、空間的自己相関が強く、正の空間的自己相関であることを意味する。また、−1 に近い値の時は、近隣の空間オブジェクトの属性値が異なるか、類似する属性値を持つ空間オブジェクトが分散しており、負の空間的自己相関であることを示す。

Moran's I の Z 値 Z_I、期待値 $E[I]$、分散 $V[I]$は、それぞれ次式から得られる。

$$Z_I = \frac{I - E[I]}{\sqrt{V[I]}}$$

$$E[I] = \frac{-1}{N-1}$$

$$V[I] = E[I^2] - E[I]^2$$

R による分析例

ここでは、関東圏（一都五県）の地価データを用いて、Moran's I を計算してみよう。データを読み込んだ後、ドロネー三角網図をもとに空間隣接行列と空間重み付け行列を作成し、Moran's I を計算する。隣接行列の行和で標準化する方法（W_W）を用いて、空間重み付け行列を適用したところ、図 5.8 に示すような結果が得られた。この結果から、地価データの Moran's I = 0.76 となり、空間的自己相関があると考えてよい。

R 分析例

```
# 地価データの読み込み
lph <- read.table("lph.csv", sep=",", header=T)
summary(lph)
```

```
# ドロネー三角網図
coords <- matrix(0, nrow=nrow(lph), ncol=2)
coords[,1] <- lph$Easting # 東経
coords[,2] <- lph$Northing # 北緯
lph.tri.nb <- tri2nb(coords)
# Moran's I
# lph$LPH が地価データを意味する
moran.test(lph$LPH, nb2listw (lph.tri.nb,style="W"))
```

この結果は、以下のようになる。

```
> moran.test(lph$LPH, nb2listw (lph.tri.nb,style="W"))

	Moran's I test under randomisation

data:  lph$LPH
weights: nb2listw(lph.tri.nb, style = "W")

Moran I statistic standard deviate = 26.9345, p-value < 2.2e-16
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
     0.7644301685     -0.0027700831      0.0008113342
```

図 5.8 Moran's I の計算結果

5.3.2 Geary's C

Geary's C は次式のように表され、0 から 2 の間の値をとる。Moran's I とは異なり、0 に近い値をとる場合は正の空間的自己相関を示し、2 に近い値をとる場合は負の空間的自己相関を示す。また、値が 1 に近いときは、空間的自己相関がなく、空間オブジェクトの属性値がランダムに分布していることを意味する。

$$Geary's\ \ C = \frac{(N-1)}{2\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}} \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(x_i - \bar{x})^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

R による分析例

Geary's C を計算したところ、図 5.9 のような結果が得られた。この結果から、地価データの Geary's C = 0.25 となり、正の空間的自己相関を示していることがわかる。

R 分析例

```
geary.test(lph$LPH, nb2listw (lph.tri.nb,style="W"))
```

```
> geary.test(lph$LPH, nb2listw (lph.tri.nb,style="W"))

	Geary's C test under randomisation

data:  lph$LPH
weights: nb2listw(lph.tri.nb, style = "W")

Geary C statistic standard deviate = 14.7215, p-value < 2.2e-16
alternative hypothesis: Expectation greater than statistic
sample estimates:
Geary C statistic       Expectation          Variance
      0.251512237       1.000000000       0.002585034
```

図 5.9 Geary's C の計算結果

5.3.3 Join count 統計量

空間オブジェクトの属性値 x_i がカテゴリカルデータの場合、Join count 統計量を用いてその空間的凝集性を表すことができる。属性値が 2 値のとき、或いは 2 値に分類できるとき、属性値を「黒（B）」または「白（W）」であるとする。このとき、隣接行列の要素 c_{ij} を用いて、属性値の組み合わせ数（count）は次のように表せる。ここで、属性値が「黒（B）」のとき $x_i = 1$ である。

$$BB = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij} x_i x_j$$

$$WW = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}(1-x_i)(1-x_j)$$

$$BW = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}\left(x_i - x_j\right)^2$$

BB、WW、BW の期待値は、それぞれ次式のようになる。

$$E(BB) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij} \cdot \left(\sum_{i=1}^{N} x_i \Big/ N\right)$$

$$E(WW) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij} - E(BB) - E(BW)$$

$$E(BW) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij} \cdot \left(\sum_{i=1}^{N} x_i \Big/ N\right) \cdot \left(\sum_{i=1}^{N} (1-x_i) \Big/ N\right)$$

このうち、BW の Join Count 統計量が期待値より低い値となり、かつ統計的で

有意であるとき、与えられた空間データが正の空間的自己相関であることを示す。

R 分析例

Join Count 統計量を計算したところ、図 5.10 のような結果が得られた。ここでは、対象地域全体の地価平均より地価が低い地区を”low”、高い地区を”high”として統計量を計算している。また、空間重み付け行列は隣接行列（W_B）をそのまま用いている。

この結果から、地価データの Join Count 統計量のうち、high:low の組み合わせとなる値は 94.0 となり、期待値と比較して低く、その Z 値は統計的に有意であることから、正の空間的自己相関を示していることがわかる。

```
lph.hi.low <- cut(lph$LPH, breaks=c(0,mean(lph$LPH),max(lph$LPH)),
labels=c("low","high"))
names(lph.hi.low) <- lph$JCODE
joincount.multi(lph.hi.low, nb2listw(lph.tri.nb, style="B"))
```

```
> joincount.multi(lph.hi.low, nb2listw(lph.tri.nb, style="B"))
          Joincount Expected Variance z-value
low:low     663.000  470.531   97.617  19.480
high:high   315.000  121.094   64.287  24.184
high:low     94.000  480.375  222.124 -25.924
Jtot         94.000  480.375  222.124 -25.924
```

図 5.10 Join Count 統計量の計算例

5.3.4 Local Moran’s I

地区 i の Local Moran’s I は次式から得られる。この指標は、Moran’s I の地区 i に関する指標であり、ローカルな空間的自己相関を意味する。Moran’s I 同様に、−1 から 1 の間の値をとる。

$$I_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2 / N} \sum_{j=1}^{N} w_{ij}(x_j - \bar{x})$$

Local Moran’s I の値と、標準化された属性値を用いて散布図を作成したとき、第一象限に分布する地区は、属性値が他の地区と比較して相対的に大きく、かつ類似する値を持つ地区が周囲にある、即ち空間的に正の自己相関をすることを意味する（図 5.11）。

総務省統計局「社会生活統計指標-都道府県の指標-」[1]に示されている、2005

年の全国の都道府県単位での糖尿病による死亡者数（人口１０万人あたり）のデータを用いて計算した Local Moran's I の分布を、図 5.12 に示す。

R 分析例

```
# データの読み込み
pref_gov <- read.table("pref_gov.txt",sep=",",
header=T,row.names=2)
pref.pnt <- readShapePoints("pref_gov.shp")
# 座標行列の定義
coords <- matrix(0,nrow(pref_gov),2)
coords[,1] <- pref_gov$X
coords[,2] <- pref_gov$Y
# ドロネー三角網図の作成
pref.tri.nb <- tri2nb(coords,
row.names=rownames(pref_gov))
# Local Moran's I の計算
LMI1 <- localmoran(pref.pnt$diabetes,
nb2listw(pref.tri.nb, style="W"))
pref.lm <- data.frame(cbind(LMI1[,1],
(pref.pnt$diabetes- mean(pref.pnt$diabetes))
/sd(pref.pnt$diabetes)),
row.names=pref.pnt$KENCODE)
colnames(pref.lm) <- c("Ii","standardized diabetes")
pref.lm
# 結果の図示
plot(pref.lm,xlab="Local Moran's I",
ylab="Standardized diabetes")
text(pref.lm,rownames(pref.lm), adj=1.2, cex=0.8)
```

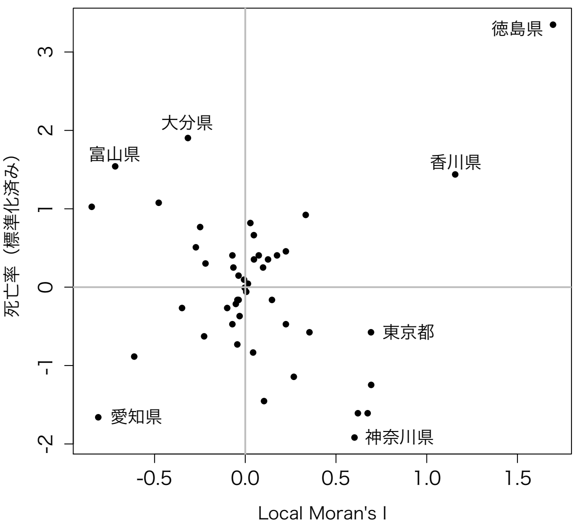


図 5.11 Local Moran's I の散布図

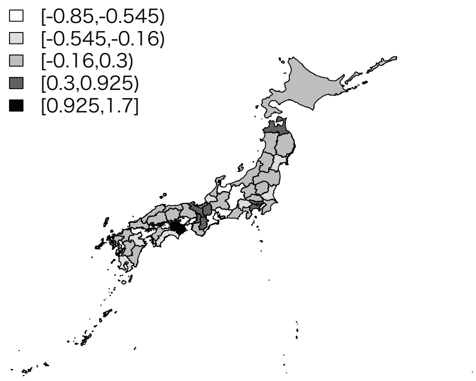


図 5.12 Local Moran's I の分布

5.3.5 G 統計量

G 統計量は、局地的な空間的自己相関を示す指標として用いられる。ホットスポットの検出などに適用されることもある。対象地域全体の G 統計量 G^*は次式のように表される。

$$G^* = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}x_i x_j}{\sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j}$$

また地区 iについて自地区を含まない G 統計量 G_iは、以下のようになる。

$$G = \frac{\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} w_{ij}x_i x_j}{\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} x_i x_j}$$

G 統計量の Z値 Z_G、期待値 $E[G]$、分散 $V[G]$は、それぞれ次式から得られる。

$$Z_G = \frac{G - E[G]}{\sqrt{V[G]}}$$

$$E[G] = \frac{-1}{N-1}$$

$$V[G] = E[G^2] - E[G]^2$$

Z値 Z_Gが高い正の値であるとき、空間的に正の自己相関をすることを意味する。全国の都道府県単位での糖尿病による死亡者数（人口１０万人あたり）のデータを用いて計算した G 統計量の結果を図 5.13 に示す。

R 分析例

G 統計量を計算したところ、以下の図 5.13 のような結果が得られた。

```
localGs.diabates <- localG(pref.pnt$diabetes,
nb2listw(include.self(pref.tri.nb), style="W"))
```

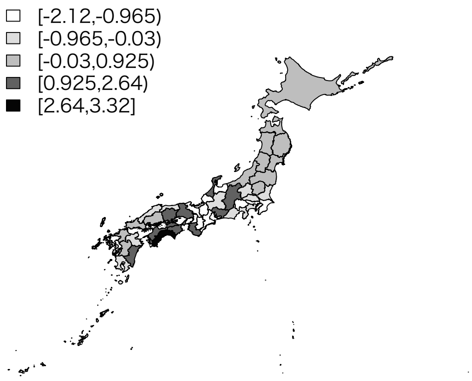


図 5.13 G 統計量の分布

参考文献

[1] 総務省統計局『社会生活統計指標-都道府県の指標-』,
http://www.stat.go.jp/data/ssds/5.htm