

第9章 空間補間

本章では、密度関数と空間補間法による空間点過程データの可視化手法を扱う。第8章では、点データ分布のランダム性を把握する方法を紹介したが、密度関数を用いれば、点データ分布の一様性や偏りを面的に可視化できる。また、連続量を属性値として持つようなマーク付き点過程について、空間補間法を用いれば、観測されない地点での属性値の予測が可能となる。空間補間法は、例えば、ボーリングデータを使った土壌成分分析や鉱石含有率の分布予測、環境観測値点データを使った気温や大気汚染物質などの分布予測などといった分野で用いられている。

まず、カーネル密度関数を用いた点密度の可視化手法について取り上げる。次に、逆距離加重法 (inverse distance weight: IDW) とクリギングという空間補間法を紹介する。クリギング補間法は、各地点の属性値に関する空間的自己相関の範囲や等方性・異方性を考慮するために、確率場データをモデル化したバリオグラム (variogram) というモデルを用いる。

9.1 カーネル密度関数

対象地域 A (面積 $|A|$) に N 個の点データ $x_i (i=1, \dots, N)$ があるとする。このとき、対象地域における点密度は $\lambda = N/|A|$ となる。地域 A の任意の地点 u における点密度の推定量 $\hat{\lambda}(u)$ は、次式のように表すことができる。

$$\hat{\lambda}(u) = \frac{1}{q(u)} \sum_{i=1}^N \frac{1}{h^2} \kappa\left(\frac{x_i - u}{h}\right)$$

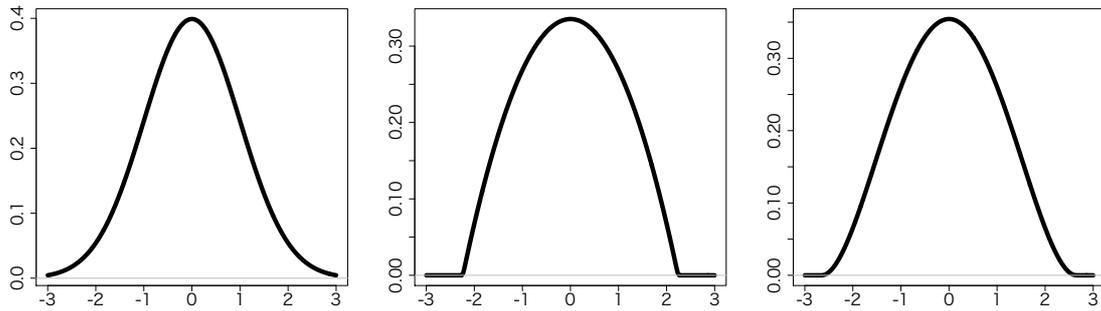
ここで、 $\kappa((x_i - u)/h)$ はカーネル関数、 h はバンド幅 (bandwidth; 帯域幅)、 $q(u)$ は境界修正を意味する。

代表的なカーネル関数として、ガウス関数 (正規分布)、イパネクニコフ関数、四次関数 (biweight 関数、quartic 関数) が挙げられる (図 9.1)。

(1) ガウス関数
$$\kappa(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$$

(2) イパネクニコフ関数
$$\kappa(u) = \frac{3}{4}(1-u^2) \cdot 1_{\{|u| \leq 1\}}$$

(3) 四次関数
$$\kappa(u) = \frac{3}{\pi}(1-u^2) \cdot 1_{\{|u| \leq 1\}}$$

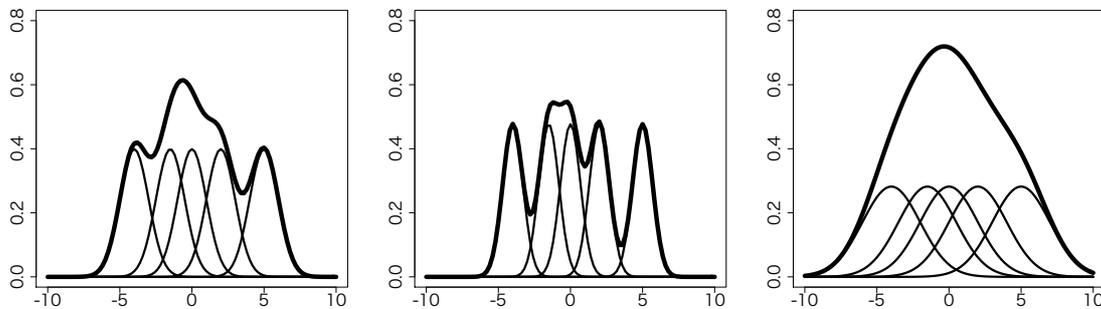


(a) ガウス関数

(b) イパネクニコフ関数

(c) 四次関数

図 9.1 カーネル関数



(a) バンド幅=1

(b) バンド幅=0.7

(c) バンド幅=2

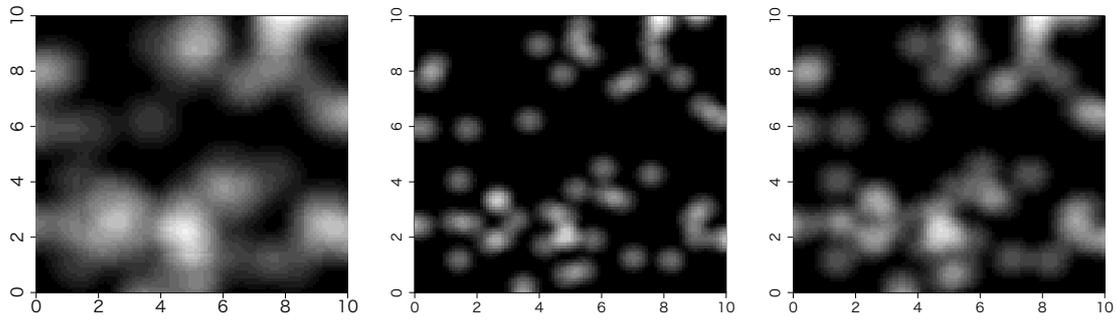
図 9.2 バンド幅を変化させた場合のカーネル密度（ガウス関数）の変化

図 9.2 ではカーネル関数にガウス関数（正規分布、図中に細い曲線で図示）を用い、バンド幅を変化させた場合に、正規分布の合成関数としてのカーネル密度関数（太い曲線）を示しているが、バンド幅の設定によって点密度が異なることがわかる。図 9.3 は、カーネル関数に四次関数を使って 2 次元空間での点密度を可視化した例を示している。

カーネル密度関数は、2 次元空間だけでなく、時間軸を含む 3 次元空間でも可視化できる。図 9.4 は、GPS をもって移動した人々の地点を、7 時から 21 時まで 1 時間毎に集約して点密度として可視化した結果を示した例である。

図 9.2 や図 9.3 に示されているように、カーネル密度の推定にはバンド幅の決定が重要な要素となる。バンド幅の決定方法には、様々な方法が提案されている。

例えば、データ数 N と標準偏差 σ または四分位範囲 IQR の小さい方 $\min\{\sigma, IQR/1.34\}$ を用いて、以下の定数を経験的に与える方法が提案されている [1]。



(a) バンド幅=1 (b) バンド幅=0.7 (c) バンド幅=0.5

図 9.3 バンド幅を変化させた場合のカーネル密度（四次関数）の変化

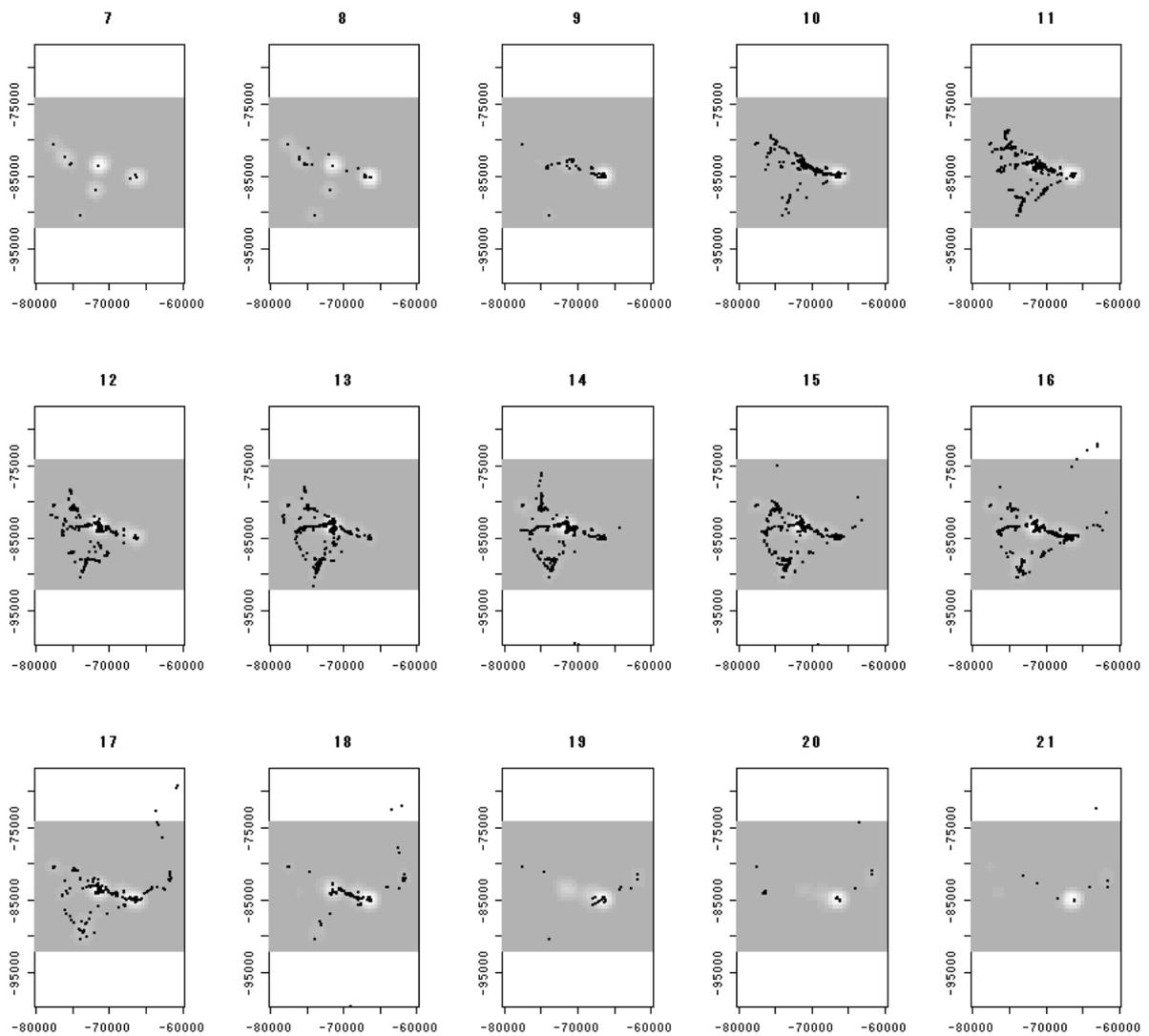


図 9.4 3次元カーネル密度の推定例

$$h = 1.06 \cdot \min\{\sigma, IQR/1.34\} \cdot N^{-0.2}$$

ガウス関数を用いる場合にはその標準偏差を用いることもある。イパネクニコフ関数は、 $\hat{\lambda}(u)$ の平均最小二乗誤差の積分を最小にするカーネル関数であることが知られている。

最適なバンド幅を与える方法として、次のような手法が提案されている。

平均最小二乗誤差 (mean square error) を最小にする方法では、 $\hat{\lambda}(u)$ の平均最小二乗誤差の積分を考え、積分値を最小にするようにバンド幅を決定する。図 9.5 には、バンド幅に応じた最小二乗誤差の計算例を示す。図中の○印は MSE が最小となるバンド幅を意味する。

マーク付き点過程を用いる場合には、交差検証対数尤度関数 (cross-validated log-likelihood function) を用いて尤度関数を最大にするバンド幅を求める方法が提案されている。図 9.6 には、バンド幅に応じた対数尤度の計算例を示す。図中の○印は、交差検証対数尤度が最大となるバンド幅を意味する。

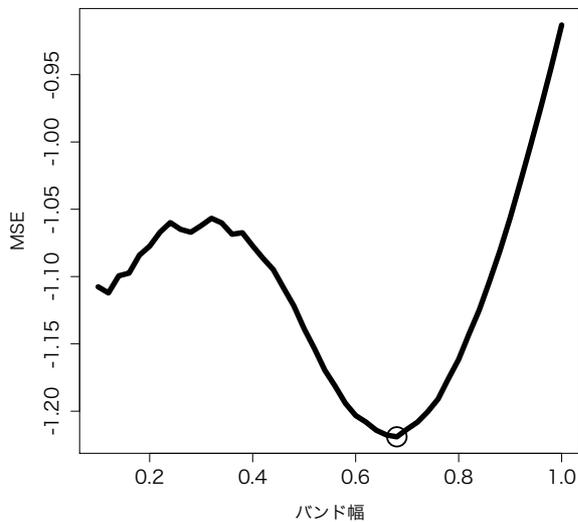


図 9.5 平均最小二乗誤差法での
バンド幅計算例

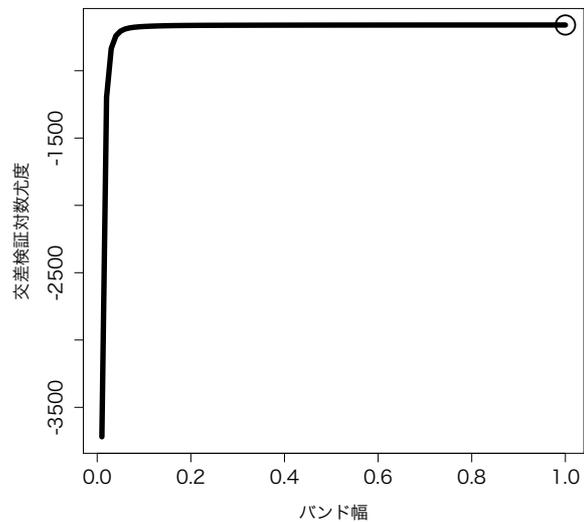


図 9.6 交差検証対数尤度関数での
バンド幅計算例

9.2 逆距離加重法

ここでは例えば、大気環境測定局で観測された大気汚染物質の濃度、ボーリング調査地点で収集された希少金属や温泉の深さ、などといったデータを分析することを考えることにしよう。

連続空間上の地点 $s_i (i=1, \dots, N)$ で、 $Z(s_i)$ というデータが観測されたとする。このとき、観測されなかった地点（測定地点がない場所） u_0 での属性値 $Z(u_0)$ を、逆距離加重法をもちいて予測したいとき、地点 u_0 と観測地点との距離による重み付け平均を用いて予測することが出来る。

$$Z(u_0) = \frac{\sum_{i=1}^N w(s_i) Z(s_i)}{\sum_{i=1}^N w(s_i)}$$

ここで $w(s_i)$ は、地点 u_0 と地点 s_i との距離関数を意味する。

$$w(s_i) = \|u_0 - s_i\|^{-p}$$

距離に対して乗するという意味で、逆距離加重法と呼ばれる

逆距離加重法を用いて予測する場合、(1) 距離関数 $w(s_i)$ を決定し、(2) 予測したい地点からどの範囲の観測地点のデータを用いて予測したいか（検索半径）を指定し、(3) 検索半径内の観測データを用いて予測地点 u_0 の属性値を予測する、という順番で予測する。



図 9.7 近隣地点の検索範囲を指定する方法

検索範囲を指定する方法は 2 通りある（図 9.7）。可変半径法（最近隣法）では、予測対象地点から最も近い観測地点を指定された地点の数だけ検索し、それらの観測地点の属性値を用いて予測する。固定半径法（固定検索半径法）では、予測対象地点を中心として、指定された半径内に入る地点を検索し、それらの値を利用して予測する。

逆距離加重法により、首都圏（一都三県）における SPM（浮遊粒子状物質）の測定局データを空間補間した例を図 9.8 に示す。

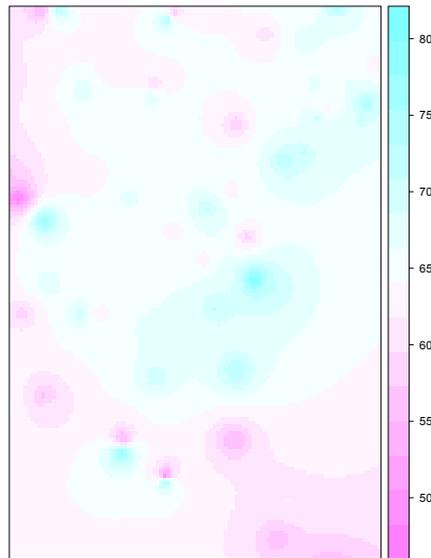


図 9.8 逆距離加重法による空間補間

R 分析例

首都圏の SPM に関する測定局ポイントデータ (`tma_spm.shp`) を使って、逆距離加重法による空間補間を適用してみよう。ここでは、`spdep` パッケージを使ってポイントデータを読み込んだ後、`gstat` パッケージの `idw()` 関数を用いて逆距離加重法を適用する例を示す。

```
# パッケージの呼び出し
library(spdep)
library(gstat)

# 測定局ポイントデータの読み込み
spm.shp <- readShapePoints("tma_spm.shp")

# 分析に必要な属性データを選びデータフレームに変換
spm <- cbind(spm.shp$ID, spm.shp$X, spm.shp$Y, spm.shp$SPM07)
colnames(spm) <- c("ID", "X", "Y", "SPM07")

spm <- as.data.frame(spm)

# 空間補間を行うメッシュデータの読み込み
mesh.grid <- read.table("mesh.csv", header=TRUE, sep=",")

# メッシュデータの位置座標を指定
```

```

coordinates(mesh.grid) <- c("X", "Y")
# メッシュデータに変換
mesh.grid <- as(mesh.grid, "SpatialPixelsDataFrame")
# 逆距離加重法を適用
spm.idw1 <- idw(SPM07*1000~1, locations=~X+Y,
data=spm, mesh.grid, idp=2)
# 結果を表示
spplot(spm.idw1["var1.pred"])

```

9.3 バリオグラム

いま、連続的な空間で、方向性によって特徴が異なることがない観測データを分析することを考えよう。観測された地点 s_i における観測データ $Z(s_i)$ が、独立した k 個の変数 $X_k(s_i)$ を用いて説明できるとする。例えば、観測データが騒音である場合、近隣の幹線道路交通量、工場や大規模小売店舗などからの距離が要因として考えられるが、これらの要因を説明変数にして騒音測定値を説明するモデルを考えるとよい。このとき、観測データ $Z(s)$ が次式のような線形回帰モデルを用いた傾向面で表されると考える。

$$Z(s) = \sum_{k=1}^m X_k(s) \beta_k + \varepsilon(s),$$

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

ただし β_k は回帰係数、 $\varepsilon(s)$ は互いに独立で同一の正規分布に従う誤差項、 σ^2 は誤差項の分散を意味する。説明変数 $X_k(s)$ の設定により、様々なモデルを推定できる。例えば $X_k(s) \equiv 1$ のとき、定数項のみのモデルとなる。説明変数に緯度経度を設定することもできる。

観測地点 ij 間の距離を $h_{ij} = \|s_i - s_j\|$ としたとき、 $Z(s_i)$ と $Z(s_j)$ の共分散 γ を h_{ij} の関数で表すことを考える。

$$2\gamma(h_{ij}) = \text{Cov}\{Z(s_i), Z(s_j)\}$$

ここから、**セミバリオグラム** を次のように定義する。セミバリオグラムと呼ばれる理由は、共分散であるバリオグラム $2\gamma(h_{ij})$ を 2 で割った値を用いていることによる。

$$\begin{aligned}\gamma(h_{ij}) &= \frac{1}{2} E \left[\left(Z(s_i) - Z(s_j) \right)^2 \right] \\ &= \frac{1}{2} E \left[\left(Z(s_i) - Z(s_i + h_{ij}) \right)^2 \right]\end{aligned}$$

さらに、地点 s_i から距離 h だけ離れた地点との関係を、次式のように表す。

$$\gamma_i(h) = \frac{1}{2} E \left[\left(Z(s_i) - Z(s_i + h) \right)^2 \right]$$

標本対すべてのバリオグラムと地点間の距離との関係を示したものを**バリオグラム雲 (variogram cloud)** という。バリオグラム雲は、[図 9.9](#) に示されるように、観測データの分布や観測値の非類似性を理解する上で有益である。

地点間の距離 h に対して階級区分 \tilde{h} 毎に細分化し、バリオグラムの平均値をとった値 $\gamma^*(\tilde{h})$ を、**標本バリオグラム (sample variogram)** という。地点間の距離が h となる組み合わせの個数を N_h とすると、標本バリオグラムは、以下のように表すことができる。

$$\gamma^*(\tilde{h}) = \frac{1}{2N_h} \sum_{i=1}^{N_h} \gamma_i(h)$$

標本バリオグラムは[図 9.10](#) に示したようになるが、標本バリオグラムに理論モデルを当てはめたものを**理論バリオグラム (theoretical variogram)** という。

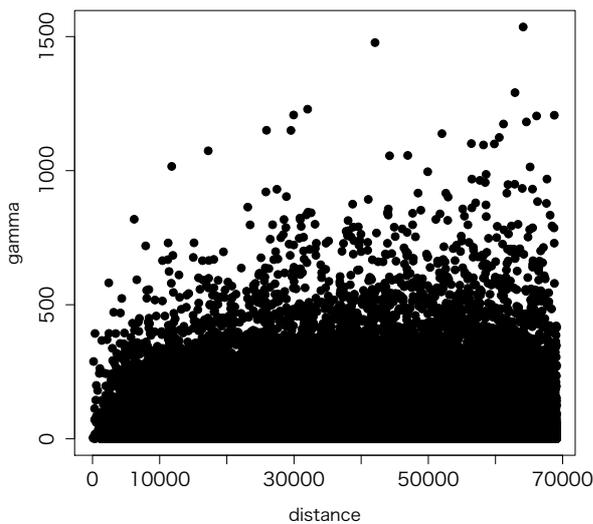


図 9.9 バリオグラム雲

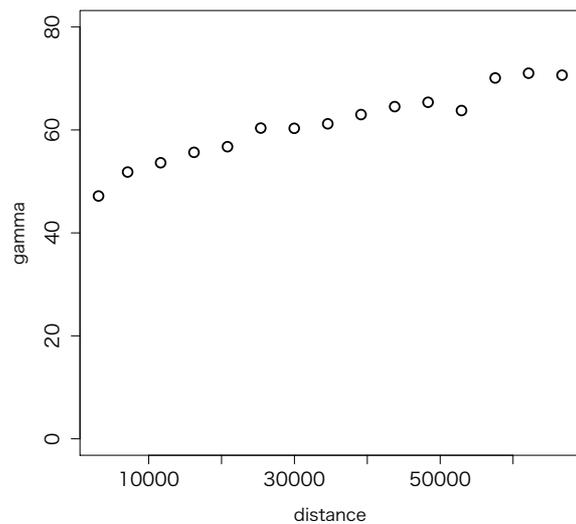


図 9.10 標本バリオグラム

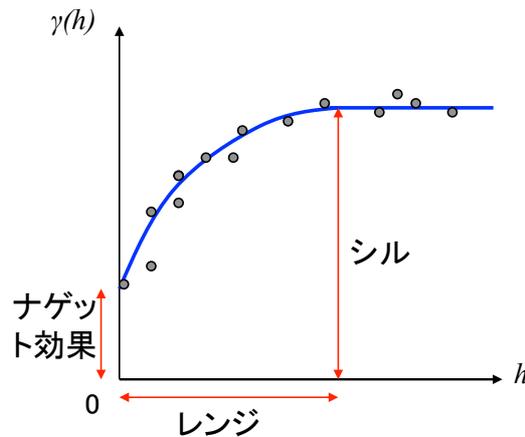


図 9.11 バリオグラムモデルのシル・レンジ・ナゲット効果

地点 s_i の属性 $Z(s_i)$ と、 s_i に対して距離 h だけ離れた地点 s_i+h の属性 $Z(s_i+h)$ との差分を考えたとき、次式で表されるセミバリエンス $\hat{\gamma}(h)$ と距離 h との関係を表したものがセミバリオグラムであると考えられることもできる。

$$\hat{\gamma}(h) = \frac{1}{2N_h} \sum_{i=1}^{N_h} (Z(s_i+h) - Z(s_i))^2$$

教科書によっては、セミバリエンスが標本バリオグラムとして紹介されていることがある¹。

一般的に、理論バリオグラムは図 9.11 のような形状をする。 $h=0$ のときのバリオグラム $\gamma(h)$ (つまり定数) をナゲット (nugget)、 $\lim_{h \rightarrow \infty} \gamma(h)$ となるときの $\gamma(h)$ をシル (sill) という。シルはデータ全体の分散も意味する。また、バリオグラムが定常状態になる距離、すなわち空間的自己相関がなくなる距離 h をレンジ (range) という。

シルはレンジに対するバリオグラムであるが、実際のデータ分析においてレンジがシルと一致することはまれなため、バリオグラムの値がシルに近づいたときの適当な値をレンジとして選択する。空間的自己相関構造がなければレンジ値を得ることはできない。

$h=0$ のとき、つまり非常に近い地点で観測されたデータであっても、観測値が

¹ セミバリエンスとセミバリオグラムの適用については、しばしば混同されるほか、用語を統一すべきなどの意見もある。

一定のばらつきをもつことがあり、それがナゲットのような定数として表現されているが、このような現象を**ナゲット効果** (nugget effect) という。同じ石の塊 ($h \approx 0$ と見なすことが出来る) から採取されたダイヤモンドや金であっても、品質が異なることがあることを考えれば、ナゲット効果の意味が理解できるだろう。

代表的なバリオグラムモデルとして、(1)指数モデル、(2)球形モデル、(3)線形モデル、(4)ガウスモデル、(5)ナゲット効果モデル、(6)Matern モデルなどが提案されている。バリオグラムモデルを推定する方法としては、①分析する人がバリオグラムを見て (見た目で) モデルの変数を設定、②通常最小二乗法や重み付き最小二乗法による推定、③最尤法や制限付き最尤推定法による推定、④ベイズ法による推定、などの方法がある [3]。

(1) 指数モデル

$$\gamma^*(h) = b \exp\left(-\frac{|h|}{a}\right) \quad (a, b > 0)$$

(2) 球形モデル

$$\gamma^*(h) = \begin{cases} b \left(1 - \frac{3|h|}{2a} + \frac{1}{2} \frac{|h|^3}{a^3}\right) & (a \geq |h| > 0) \\ b + a & (|h| > a) \\ 0 & (|h| = 0) \end{cases}$$

(3) 線形モデル

$$\gamma^*(h) = \begin{cases} b + a|h| & (|h| > 0, a \geq 0) \\ 0 & (|h| = 0) \end{cases}$$

(4) ガウスモデル

$$\gamma^*(h) = \begin{cases} b - c \cdot \exp\left(-\frac{|h|^2}{a^2}\right) & (|h| > 0, a > 0) \\ 0 & (|h| = 0) \end{cases}$$

(5) ナゲット効果モデル

$$\gamma^*(h) = \begin{cases} b & (|h| > 0, b \geq 0) \\ 0 & (|h| = 0) \end{cases}$$

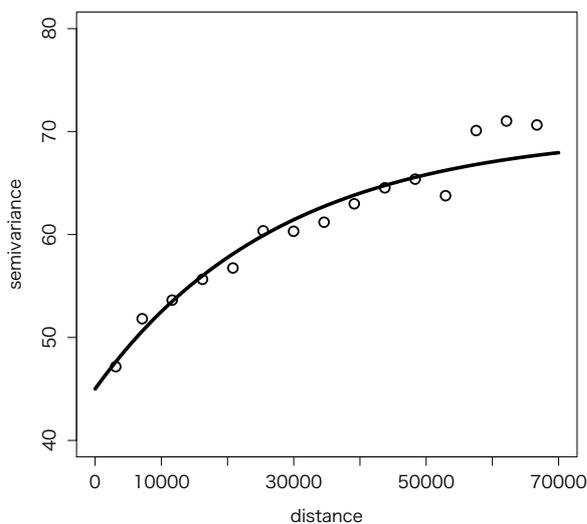
(6) Matern モデル

$$\gamma^*(h) = \begin{cases} b + c \left(1 - \frac{1}{2^{\nu-1}} \Gamma(\nu) \left(\frac{|h|}{a} \right)^{\nu} K_{\nu} \left(\frac{|h|}{a} \right) \right) & (|h| > 0, a > 0) \\ 0 & (|h| = 0) \end{cases}$$

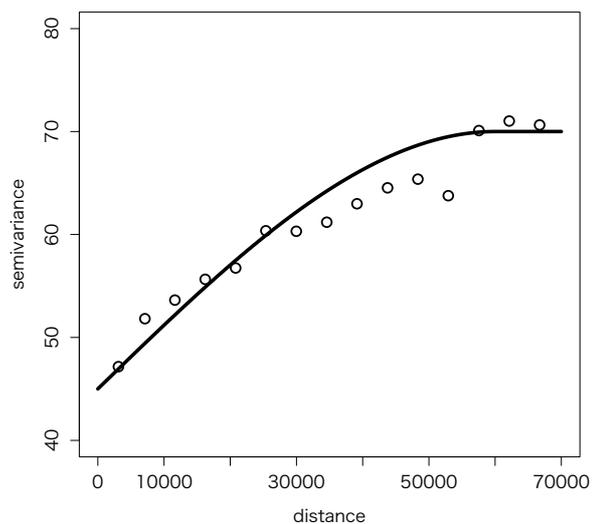
ただし K_{ν} は修正 Bessel 関数である。

上述の 6 つのバリオグラムモデルの推定例を、[図 9.12](#) に示す。図中の \circ は経験バリオグラムである。ここで、これらのバリオグラムモデルを計算する際に、緯度経度による傾向面を説明変数としたモデルを採用している。

このうち、バリオグラムモデルに球形モデルを選び、 (N_j/h_j^2) で重み付けした) 重み付け最小二乗法 (WLS)、通常最小二乗法 (OLS)、制限付き最尤法 (REML) を用いてナゲット効果、シル、レンジの各パラメータを推定した結果を、[図 9.14](#) に示す。図中の \circ は経験バリオグラムである。

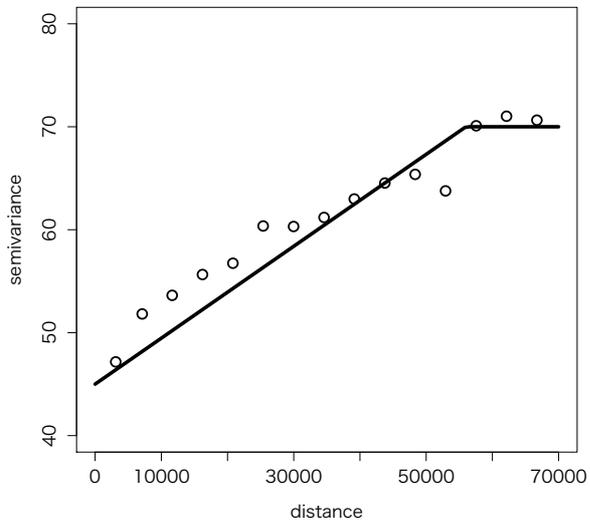


(1) 指数モデル

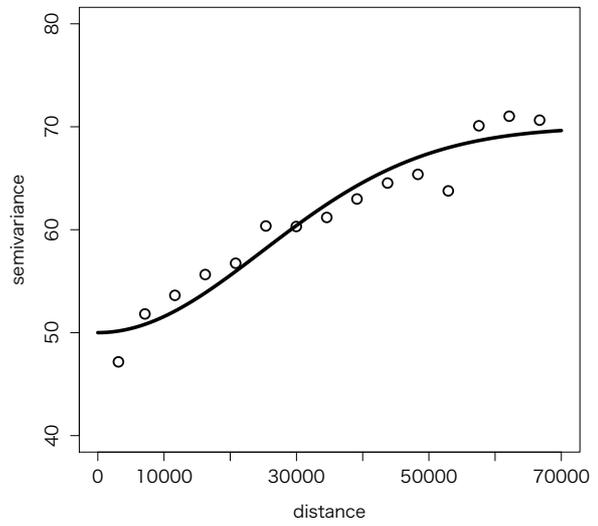


(2) 球形モデル

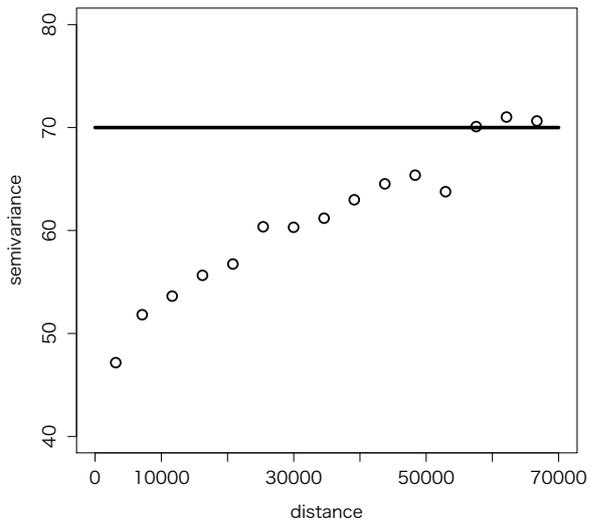
図 9.12 バリオグラムモデル



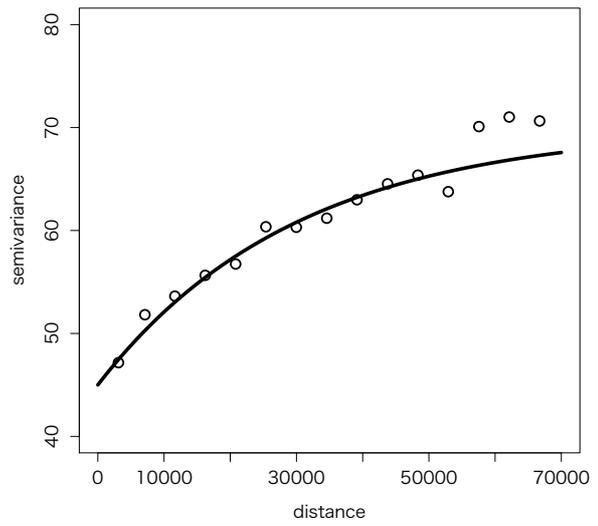
(3) 線形モデル



(4) ガウスモデル



(5) ナゲット効果モデル



(6) Matern モデル

図 9.12 バリオグラムモデル (つづき)

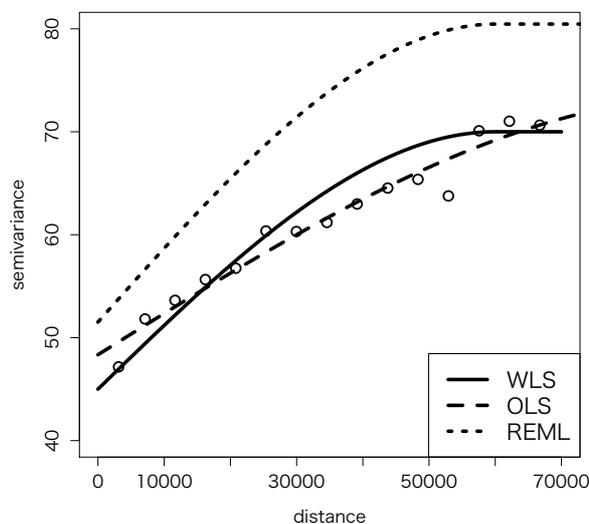


図 9.13 推定法によるバリオグラムモデルの違い

R 分析例

9.2 の分析例で用いたデータを使って、バリオグラムモデルを求めよう。すでに `gstat` パッケージが呼び出され、データ `spm` が読み込まれているものとする。

```
# 緯度経度の指定
coordinates(spm) <- c("X","Y")

# 緯度経度によるトレンドを説明変数とするバリオグラム
spm.var1 <- variogram(SPM07*1000~X+Y, data=spm)

# バリオグラムの表示
plot(spm.var1)

# バリオグラム雲の表示
plot(variogram(SPM07*1000~X+Y, data=spm, cloud=TRUE))

# バリオグラムモデル (指数モデル)
spm.model1 <- vgm(psill=25, model="Exp", range=28000, nugget=45)
plot(spm.var1, spm.model1, cex=1.5, lwd=4)

# バリオグラムモデル (球形モデル)
spm.model2 <- vgm(psill=25, model="Sph", range=60000, nugget=45)
plot(spm.var1, spm.model2, cex=1.5, lwd=4)
```

```

# 異なる方法でのバリオグラムモデルの当てはめ（球形モデル）
# 重み付け最小二乗法（WLS）
fit.variogram(spm.var1, spm.model2, fit.method=7)
# 通常最小二乗法（OLS）
fit.variogram(spm.var1, spm.model2, fit.method=6)
# 制限付き最尤法（REML）
fit.variogram.reml(SPM07*1000~X+Y,data=spm,
model=vgm(25, "Sph", 60000, 45))

```

異なる方法でバリオグラムモデルを当てはめた結果は、以下のようになる。

```

> fit.variogram(spm.var1, spm.model2, fit.method=7)
model  psill  range
1  Nug 46.22705  0.00
2  Sph 19.98820 52101.86
> fit.variogram(spm.var1, spm.model2, fit.method=6)
model  psill  range
1  Nug 48.34058  0.00
2  Sph 25.15494 93784.44
> fit.variogram.reml(SPM07*1000~X+Y,data=spm, model=vgm(25, "Sph", 60000, 45))
model  psill range
1  Nug 51.51024  0
2  Sph 28.94858 60000

```

重み付け最小二乗法の推定結果を見ると、psill の列に Nug=46.22705、Sph=19.98820 となっている。これは、ナゲット効果が 46.22705 であり、シルは $46.22705 + 19.98820 = 66.21525$ であることを意味している。またレンジは 52101.86 となっている。

9.3.2 異方性

実際のデータを用いてバリオグラムを推定するとき、方向によってバリオグラムが異なった特徴を示すことがある。このことを**異方性**（anisotropy）という。異方性を考慮したバリオグラムモデルを推定するために、(1)（北からみた）主軸の方位角（angle）、(2)主軸（方位角）に対する許容範囲（tolerance）、(3)バンド幅（bandwidth）、(4)主軸の拡大率（ratio）、などの変数を設定する[2]（[図 9.14](#)）。

方向別にセミバリオグラムのシルが変化せず、レンジが変化するような場合を**幾何学的異方性**（geometric anisotropy）といい、方向別にシルが変化する場合を**帯状異方性**（zonal anisotropy）という。90° 毎に異方性を考慮したバリオグラムの推定例を [図 9.15](#) に示す。

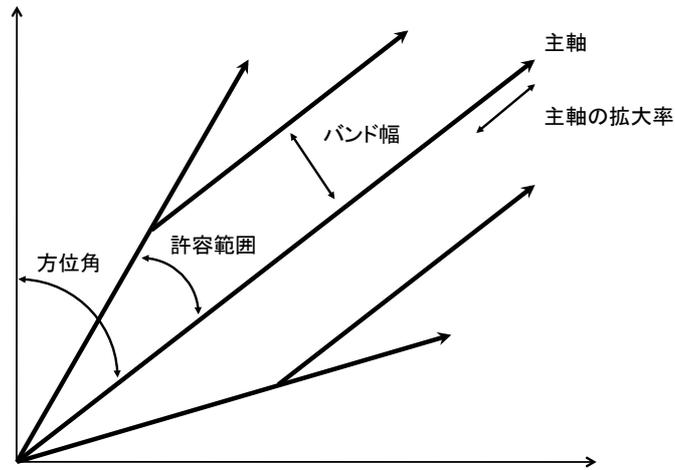


図 9.14 異方性バリオグラムモデルで設定する変数

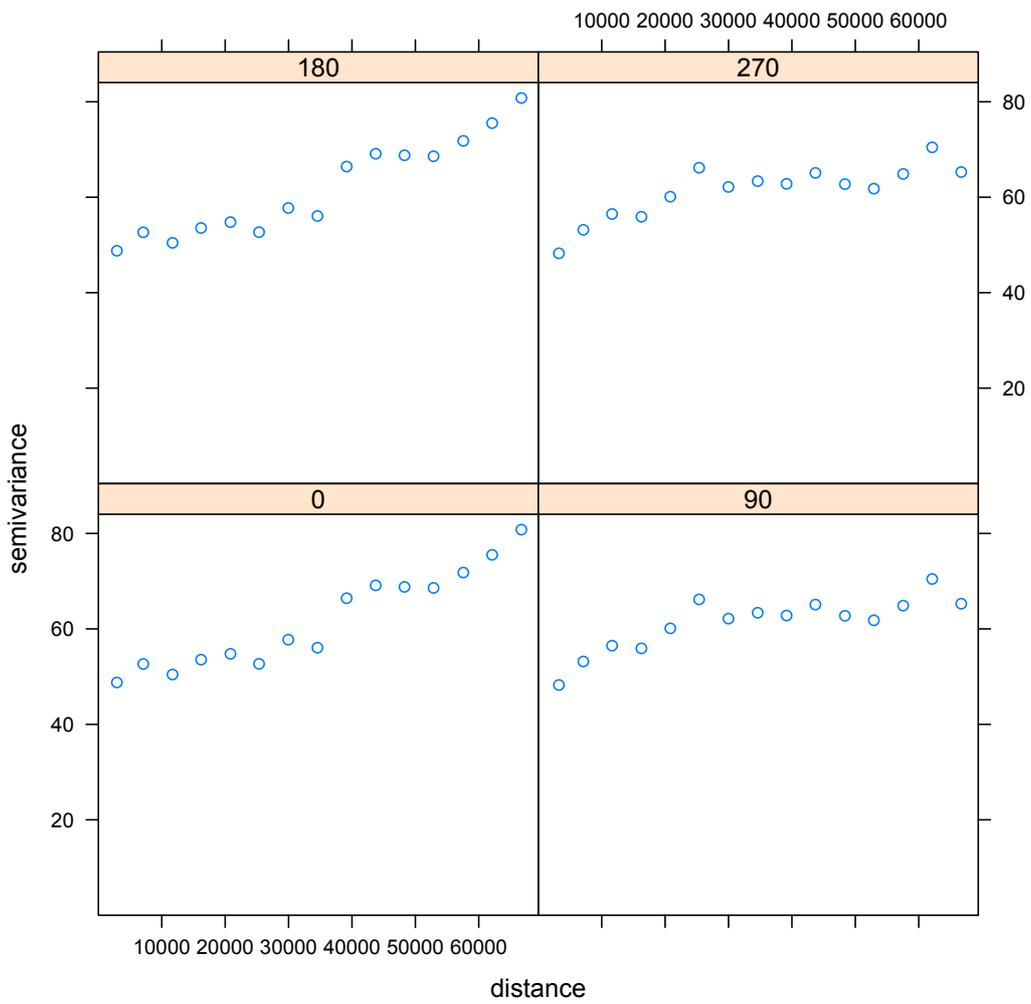


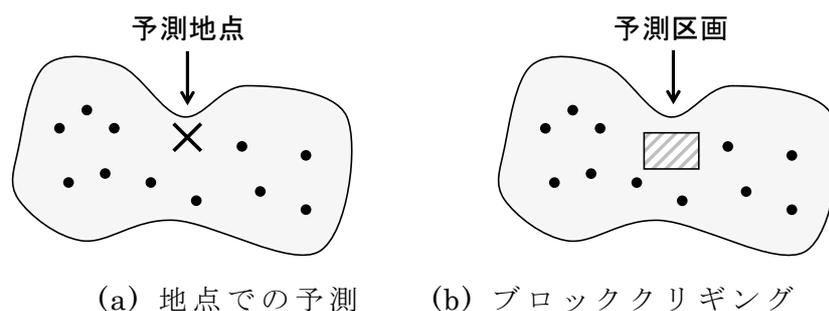
図 9.15 異方性バリオグラムモデルの推定例

9.4 クリギング

9.4.1 クリギングの考え方

クリギングにより空間補間を行う場合、予測手法には大きく二つの考え方がある[4] (図 9.16)。

一つは、観測データ $Z(s_i)$ が与えられたときに、観測されない地点 u_0 での属性値 $Z^*(u_0)$ をピンポイントで予測する方法である。単純型クリギング (simple kriging)、通常型クリギング (ordinary kriging)、普遍型クリギング (universal kriging) などの手法がこの方法に含まれる。二つ目は、ある領域内での平均的な属性値を予測する方法である。この方法はブロッククリギング (block kriging) などと呼ばれる。予測する領域は、方形だけでなく任意のポリゴンなどを指定することができる。



(a) 地点での予測 (b) ブロッククリギング

図 9.16 クリギング補間による予測手法

代表的なクリギング法に、通常型クリギングが挙げられる。通常型クリギングでは、地点 u_0 での属性値 $Z^*(u_0)$ を予測するときに、観測されたデータ $Z(s_i)$ と重み係数 w_i を用いて、重み付け平均として予測する。

$$Z^*(u_0) = \sum_{i=1}^N w_i Z(s_i),$$
$$\sum_{i=1}^N w_i = 1$$

このとき、次式で表される予測誤差 $\varepsilon(u_0)$ を最小にするような重み係数の組み合わせを計算することが要求される。

$$\varepsilon(u_0) = Z^*(u_0) - \sum_{i=1}^N w_i Z(s_i)$$

このことは、平均二乗予測誤差 $\sigma^2(u_0)$ を最小にすることにほかならない。平均二乗誤差を計算すると、次式が得られる。

$$\sigma^2(u_0) = -\sum_{i=1}^N \sum_{j=1}^N w_i w_j \gamma(h_{ij}) + 2 \sum_{i=1}^N w_i \gamma(h_{i,u_0})$$

ここで、 $\gamma(h_{ij})$ および $\gamma(h_{i,u_0})$ はバリオグラムである。つまり、通常型クリギングでは、バリオグラムを用いて重み付け係数を求めていることになる。

バリオグラムモデルが得られているときには、観測地点同士のバリオグラム $\gamma(h_{ij})$ だけでなく、予測地点とのバリオグラム $\gamma(h_{i,u_0})$ も既知である。したがって、平均最小二乗誤差を最小にするような重み付け係数の組み合わせは、ラグランジュ乗数 λ を用いて求めることができる。

単純にするために、3つの観測地点 ($i=1,2,3$) から地点 u_0 の値を予測することを考えると、重み付け係数とラグランジュ乗数を未知変数とする目的関数 $f(w_1, w_2, w_3, \lambda)$ は次式のようになる。

$$f(w_1, w_2, w_3, \lambda) = -\sum_{i=1}^3 \sum_{j=1}^3 w_i w_j \gamma(h_{ij}) + 2 \sum_{i=1}^3 w_i \gamma(h_{i,u_0}) - 2\lambda \left(\sum_{i=1}^3 w_i - 1 \right)$$

この式を、重み付け係数とラグランジュ乗数について偏微分することにより、以下のような連立方程式が得られる。この連立方程式を解くことにより、重み付け係数とラグランジュ乗数を求めることができる。

$$\begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \gamma(h_{13}) & 1 \\ \gamma(h_{21}) & \gamma(h_{22}) & \gamma(h_{23}) & 1 \\ \gamma(h_{31}) & \gamma(h_{32}) & \gamma(h_{33}) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(h_{1u_0}) \\ \gamma(h_{2u_0}) \\ \gamma(h_{3u_0}) \\ 0 \end{bmatrix}$$

9.4.2 通常型・単純型・普遍型クリギング

前項で紹介した通常型クリギングを基本として、いくつかの重要なクリギング補間法を導くことができる。9.3.1 でバリオグラムモデルを紹介した際に、観測データが次式のような傾向面で予測されることを述べた。

$$Z(s) = \sum_{k=1}^m X_k(s) \beta_k + \varepsilon(s)$$

この性質を用いれば、未知パラメータ β_k に関する不偏推定量 $\hat{\beta}_k$ が得られたとき、地点 u_0 での不偏推定量 $\hat{Z}(u_0)$ は次式のように表すことができる。

$$\hat{Z}(u_0) = \sum_{k=1}^m X_k(u_0) \hat{\beta}_k + \varepsilon(u_0)$$

$\hat{Z}(u_0)$ と $Z^*(u_0)$ の普遍性を確保すると、

$$E[\hat{Z}(u_0) - Z^*(u_0)] = 0$$

となる。すると、次の2式を得ることができる。

$$\sum_{k=1}^m X_k(u_0) \hat{\beta}_k - \sum_{i=1}^N w_i Z(s_i) = 0$$

$$\sum_{k=1}^m \hat{\beta}_k \left(x(u_0) - \sum_{i=1}^N w_i x(s_i) \right) = 0$$

ここで、 $\hat{\beta}_k \neq 0$ であることから、次式で表される普遍性条件が得られる。

$$x(u_0) - \sum_{i=1}^N w_i x(s_i) = 0$$

このモデルについて、平均二乗予測誤差 $\sigma^2(u_0)$ を最小にするような重み付け係数の組み合わせを求めることにより、普遍型クリギングの推定結果が得られる。

図 9.17 に、普遍型クリギングを用いた空間補間の例を示す。

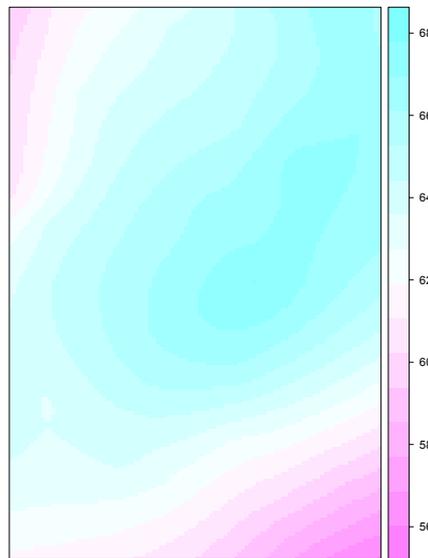


図 9.17 普遍型クリギングによる空間補間

ちなみに、 $m=1$ かつ $x_1(u_0) \equiv 1$ としたとき、通常型クリギングとなる。

さらに、何らかの理由で不偏推定量 $\hat{\beta}_k$ が既知であるとき、これを用いて得られる予測結果が単純型クリギングとなる。

予測対象となる観測データに対して、すべての観測地点で相関のある他の観測データが得られているとする。このとき、相関のある観測データを説明変数として予測対象となるデータを予測する方法を、コクリギング (cokriging; 共クリギング、共変量クリギング) という。

クリギングを用いた予測には、様々な仮定を置いているため、予測誤差が実際の推定誤差と乖離する場合も生じる。そこで、モデルの適合度を検証するために、交差検証により実際の推定誤差とモデルによる予測誤差の大きさを比較する方法なども用いられている。

R 分析例

9.2 及び 9.3 の分析例に示した手順により、バリオグラムモデルが既に得られているとする。このとき、`gstat` パッケージの `gstat()` 関数を用いてクリギング補間に必要な情報を定義し、`predict()` 関数によりクリギング補間を適用する。

ここでは、バリオグラムモデルとしてガウスモデル(9.3 分析例の `spm.model1`) を用い、普遍クリギングを適用している。

```
# クリギング補間に必要な情報を定義
spm.gu <- gstat(id="ID", formula=SPM07*1000~X+Y, data=spm,
model=spm.model1)
# クリギング補間 (普遍クリギング)
spm.pu <- predict(spm.gu, mesh.grid)
# 結果の表示
spplot(spm.pu[1])
```

9.4.3 ベイズ・クリギング

これまで紹介した方法では、まずバリオグラムモデルを平均最小二乗誤差法により推定し、その次にバリオグラムを用いてラグランジュ乗数法によりクリギング補間を行うという手順であった。そのため、クリギング補間を行う際には、バリオグラムのもつ不確実性を無視して、バリオグラムモデルのパラメータが既知であることが前提となっていた。

近年、ベイズ推定法を用いて、バリオグラムモデルのパラメータが未知である場合でも、バリオグラムモデルの不確実性を考慮してクリギング補間を行うベ

ズ・クリギング (Bayesian Kriging) が提案されている。ベイズ推定法を用いれば、モデル推定上の仮定を緩和して柔軟なモデリングを行えることが期待できる。ベイズ・クリギングは、これまで教科書などであまり紹介されることがなかったが、有効な手法の一つと考えられるため、ここで取り上げておきたい。

クリギング補間を行う上で未知パラメータとなっている主要な変数は、観測データの傾向面を表す線形回帰モデルの回帰係数 β 、バリオグラムモデルのシル b (σ^2)、レンジ a (ϕ)、ナゲット c (τ^2) である。バリオグラムモデルに Matern モデルを採用する際には、その円滑化要素 ν も含まれる。

観測地点データ $Z(s)$ が得られたとき、 $\theta = \{\beta, a, b, c\}$ に関する事前情報を与えると、地点 u_0 での予測値 $Z(u_0)$ の事後情報は、次式で与えられる。

$$p(Z(u_0)|Z(s)) = \int_{\theta \in \Theta} p(Z(u_0)|\theta, Z(s)) p(\theta|Z(s)) d\theta$$

従って、ベイズの定理より、 θ は以下のようにして得られる。

$$p(\theta|Z(s)) = \frac{p(Z(u_0)|\theta, Z(s)) p(\theta|Z(s))}{\int_{\theta \in \Theta} p(Z(u_0)|\theta, Z(s)) p(\theta|Z(s)) d\theta}$$

このことから、 θ に関する事前情報を与えることで、マルコフ連鎖モンテカルロ法により事後情報をベイズ推定することにより、未知パラメータを計算することができる。

ところで、クリギングモデルの側から未知パラメータのベイズ推定を考えると、以下のようなことがわかる。

いま、観測地点データ $Z(s)$ が与えられたときの予測地点 u_0 での予測値 $Z(u_0)$ の事後分布が独立な正規分布に従うとする。潜在的な空間過程に関する変数 $T_k(u_0)$ を導入することにより、次式のような関係が得られる。

$$Z(u_0)|Z(s) \sim N\left(\sum_{k=1}^m X_k(u_0)\beta_k + \sum_{k=1}^m \sigma_k T_k(u_0), \tau^2\right)$$

するとこのとき、クリギングモデルは以下のような線形混合モデルとして表すことができる。

$$\begin{aligned} Z(u_0) &= \sum_{k=1}^m X_k(u_0)\beta_k + \sum_{k=1}^m \sigma_k T_k(u_0) + \varepsilon(u_0) \\ T_k(u_0) &\sim N(0, \text{Cor}_k(\phi_k)) \\ \varepsilon(u_0) &\sim N(0, \tau^2 I) \end{aligned}$$

ただし、 $\sum_{k=1}^m \sigma_k T_k(u_0)$ は平均 0 分散 σ^2 のガウス定常過程、 $Cor_k(\phi_k)$ は適当なパラメータ

ϕ_k で表される地点間の相関行列である。

得られた未知パラメータのうち σ^2 、 τ^2 、 ϕ は、それぞれバリオグラムモデルの未知パラメータのシル、ナゲット、レンジに対応することが知られている [5]。

参考文献

- [1] Silverman, B. W. (1986) *Density Estimation*, Chapman & Hall.
- [2] 小山修平・橋淳治訳 (1993) 『GIS の応用-地域系・生物系環境科学へのアプローチ』, 森北出版. (Johnston, C.A. 原著 (1998) *Geographic Information Systems in Ecology*, Blackwell)
- [3] Diggle, P.J. and P.J. Ribeiro Jr. (2007) *Model-based Geostatistics*, Springer.
- [4] 地球統計学研究委員会訳編/青木謙治監訳 (2003) 『地球統計学』, 森北出版. (Wackernagel, H. (1995) *Multivariate Geostatistics*, Springer.)
- [5] Cressie, N. (1993) *Statistics for Spatial Data*, Wiley.