Echelon 解析における lattice データ分析

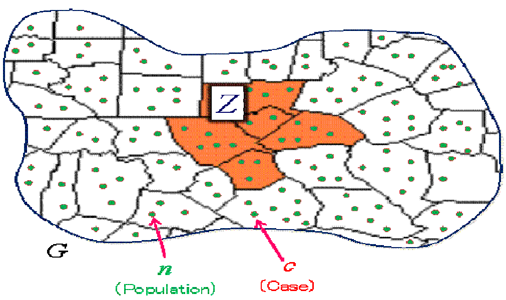
豆野貴之（岡山大学大学院環境学研究科）
栗原考次（岡山大学大学院環境学研究科）

1. はじめに

Echelon 解析(Kurihara et al,2004)は配列上に分けられた地図上の 1 変量データに対して、空間的な位置を表面上のデータ高低（濃淡）に基づき分割し、空間データの位相的な構造を系統的かつ客観的に見つけるために開発された解析法である。lattice データとは、そのリモートセンシングなどで得られた n×m の格子データのことである。本講演では lattice データでの Echelon 解析の解析法と、lattice データ分析の際、空間スキャン統計量(Kulldolff,1997)の計算に使用される空間領域の検出法について述べる。

2. 空間スキャン統計量

領域毎に得られるデータにおいて有意に高い値（ホットスポット）を見つけるために用いる。すべての領域を G とし、G の部分集合を Z とする。**N(G)**をすべての領域 G での母集団の数、**N(Z)**を領域 Z 内の母集団の数とする。**C(G)**をすべての領域 G での属性をもつものの数、**C(Z)**を領域 Z 内の属性をもつものの数とする。領域 Z の内外の確率は以下に示す。



$$\begin{cases} p_1 = \dfrac{c(Z)}{n(Z)} \\ p_2 = \dfrac{c(Z^c)}{n(Z^c)} = \dfrac{c(G)-c(Z)}{n(G)-n(Z)} \end{cases}$$

帰無仮説 $H0$: $p1 = p2 = p$ に対して
対立仮説 $H1$: $p1 > p2$ を立てる。
ここでは、ポアソン分布に基づくモデルを考える． 全領域 G で属性をもつ数が $c(G)$ になる確率は、

$$\frac{\exp[-p_1 n(Z) - p_2(n(G)-n(Z))][p_1 n(Z) + p_2(n(G)-n(Z))]^{c(G)}}{c(G)\,!}$$

であり領域 G での地点 x での密度は

$$\begin{cases} \dfrac{p_1 n(x)}{p_1 n(Z) + p_2(n(G)-n(Z))} & if \ \ x \in Z \\ \dfrac{p_2 n(x)}{p_1 n(Z) + p_2(n(G)-n(Z))} & if \ \ x \notin Z \end{cases}$$

領域 Z を与えた下での尤度関数 L(Z)は

$$L(Z) = \frac{\exp[-c(G)]}{c(G)\,!}\left(\frac{c(Z)}{n(Z)}\right)^{c(Z)}\left(\frac{c(G)-c(Z)}{n(G)-n(Z)}\right)^{c(G)-c(Z)}\prod_{x_i}^{n} n(x_i)$$

ホットスポットは尤度比λ(Z)が最大となる連続する領域 Z である。

$$\lambda(Z) = \frac{\underset{Z}{Max}\,L(Z)}{L_0} = \frac{\left(\dfrac{c(Z)}{n(Z)}\right)^{c(Z)}\left(\dfrac{c(G)-c(Z)}{n(G)-n(Z)}\right)^{c(G)-c(Z)}}{\left(\dfrac{c(G)}{n(G)}\right)^{c(G)}}$$

$$= \left(\frac{c(Z)}{e(Z)}\right)^{c(Z)}\left(\frac{c(G)-c(Z)}{e(G)-e(Z)}\right)^{c(G)-c(Z)}$$

ここで L_0 は帰無仮説　$p1 = p2 = p$ の下での尤度関数である。

$$L_0 = \sup_p \frac{\exp[-pn(G)]}{c(G)\,!} p^{c(G)} \prod_{x_i}^{n} n(x_i)$$

$$= \frac{\exp[-c(G)]}{c(G)\,!} \left(\frac{c(G)}{n(G)}\right)^{c(G)} \prod_{x_i}^{n} n(x_i)$$

また $e(Z)$は、領域 Z 内での属性の期待値であり、$e(G)=c(G)$である。

領域 G 内において対数尤度比統計量 logλ(Z)が高い値をとるような領域 Z が尤度の高いホットスポット領域と言える。

3. Echelon 解析

空間スキャン統計量を利用した新たなホットスポット検出法として我々は Echelon 解析を利用した Echelon Scan を提唱している。Echelon 解析とは空間データに対して、空間的な位置を表面上のデータ高低（濃淡）に基づき位相的に同じ領域（階層; echelon）に分割する解析手法。空間データの持つ位相的な構造を系統的かつ客観的に表現する。

2	24	8	15	3
10	1	14	22	5
4	13	19	23	25
20	21	12	11	17
16	6	9	18	7

図１

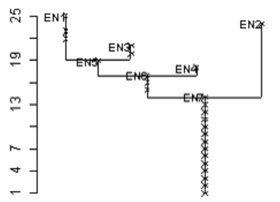


図２

図１は５×５の lattice データ,解析すると図２が得られ、それぞれ echelon 値が高い順にデンドログラムに表示される。

4. ポリオミノ理論

パターンＺを検出する際に効果的な理論がある。ポリオミノ（polyomino）P とは、平面状で頂点が整数座標をもつ単位正方形の有限和でつくられる図形で連結なもののことである。底辺から順番に一段ずつ「行」を積み重ねることで HC-ポリオミノを構成するものとする。i 段目に r 個の正方形があるとき、その上の i+1 段目に s 個の正方形をつけ加えようとすれば、r+s-1 通りの重ね方が可能である。よって

$$f(n) = \sum (n_1 + n_2 - 1)(n_2 + n_3 - 1) \cdots (n_s + n_{s+1} - 1)$$

が得られる。

ただしここで和は 2^{n-1} 通りある n 組成 $n = n_1 + n_2 + \cdots + n_{s+1}$ のすべてをわたって加える。

５．最後に

本論文ではポリオミノ理論を用いて計算機で n=18 まで(lattice データでは 6×6 行列)を計算させた、回転なし 1540820542 個、回転あり 192622052 個を検出するのに９時間を要し、領域 Z のパターンの量の膨大さがわかり、改良の必要性もある。空間スキャン統計量を求めるにはこの隣接パターンを組み合わせて、n×m のブロックにあてはめて考えることが必要である。

参考文献

栗原　考次　リモートセンシングデータ及び空間データの構造分析に関する研究

リチャードスタンレイ　数え上げ組み合わせ論 I

Kulldorff M.(1997).A spatial scan statistic,CommunicationsinStatistic,Theory and Methods

Kurihara,K.Myers,W.L.,andPatil,G.P.(2000).Echelon analysis of the relationship between population and land cover patterns based on remote sensing data.Community Ecology