

# エシロン解析に基づく 都心地区への都市間旅客来訪地点の分析

一万田 凌<sup>1</sup>・塚井 誠人<sup>2</sup>

<sup>1</sup>学生会員 広島大学 大学院先進理工系科学研究科 (〒739-8527 広島県東広島市鏡山 1-4-1)  
E-mail: m216915@hiroshima-u.ac.jp

<sup>2</sup>正会員 広島大学准教授 先進理工系科学研究科 (〒739-8527 広島県東広島市鏡山 1-4-1)  
E-mail: mtukai@hiroshima-u.ac.jp

COVID-19 による緊急事態宣言や蔓延防止等重点措置などの社会生活の制限が続いている。都心の人流の減少は、短期的には経済活動の停滞を招き、中長期的には人流と都市施設配置の間でのミスマッチによる不効率が生じる。このミスマッチが都市施設配置の再編に結びつくか否かは、COVID-19 からの回復の速さに依存するが、計画的な土地利用を促す上では、再編の好機とも考えられる。しかし計画的な再編実施には、まずミスマッチを起こす地点や施設の同定が必要である。そのような手法として、Hot spot 検出手法の対象地域全体への適用が考えられるが、実証的な検討が不足している。本稿では既存の Hot spot 検出手法の整理を行って、都市域全体を網羅する人流と都市施設配置のミスマッチを同定する手法を検討する。まずは実証分析にあたって留意すべき点を整理し、分析手順を定める。

**Key Words:** 空間スキャン検定, 都市施設, 人流データ

## 1. はじめに

COVID-19 による緊急事態宣言や蔓延防止等重点措置などの社会生活の制限が続いている。外出の自粛により、市中の人流は COVID-19 以前よりも減少した状態が継続している。人流の減少は都市内各地区で不均一に起こっており、たとえば利用者や行政が感染を懸念する施設が多く分布する地区では人流の減少が著しい一方で、リモートワーク等の影響で住宅地では日中の在宅者が多くなるなどの変化が現れている。都心の人流の減少は短期的には経済活動の停滞を招くほか、中長期的には人流と都市施設配置の間のミスマッチにより、施設の混雑や有休などの不効率が生じる。このミスマッチが市場原理による都市施設配置の再編に繋がるか否かは、COVID-19 からの回復の速さに依存するが、計画的な都市空間利用を進めるうえでは好機とも考えられる。しかしこれは一部の代表的地点の人出のモニタリングでは不可能である。ミスマッチを起こす地点や施設を同定するには、都市域全体を網羅した Hot spot 検出手法の開発が必要である。

本稿では、都市域全体を網羅した人流と都市施設配置のミスマッチを明らかにする手法について、既存手法を参照した分析手順を整理して、実証分析を行う。

## 2. Hotspot 検出手法の概要

Hot spot は、地理情報に紐づく属性値（たとえば時間帯別の人流や都市基盤施設数など）が他地点と比較して顕著に高い地点、または近傍を含む領域を指す。Hot spot 検出問題は、連続空間上に分布する点にも適用できるが、その方法論は離散空間上の領域の場合とは異なる。本稿ではポイント情報ではなく、ポリゴンやメッシュなどを対象とする Hot spot 検出手法を検討する。

Hot spot 候補地点の抽出は、たとえば地理情報を無視して属性値の確率密度を算出し、パーセンタイル基準で抽出の閾値を定めた上で、それを超過する地点の集合を定める（サンプルを昇順に並べて下からパーセンタイルを増分する場合）等の手順によればよい。問題は抽出した Hot spot 候補点が統計的に有意なことを同定する手順にあり、以下に示す 2 点の課題がある。一つは空間的な集塊性をどう判断するか（課題 a）、もう一つは他地点と比較して、統計的観点から有意に属性値が高いことをどう判断するか（課題 b）、である。

対象空間内の部分領域が比較的大きく、部分領域自体の空間結合（この場合は、隣接する部分領域の結合）を議論する余地がない場合は、部分領域単位の空間集塊性

表 1 エシエロン解析アルゴリズムの概要

指標を用いればよい。代表的な指標は、Criff and Ord による Local Index of Spatial agglomeration (LISA) や、Local Moran's I などである。しかし後述するモバイル空間統計などでは、4 次メッシュ単位の人出データが観測されており、データが得られる最小の部分領域よりも大きな領域に中心市街地が広がっている。この場合は、上述した 2 点への対応が必要になる。

ここから、栗原・石岡の提案するエシエロン解析に基づく手法を検討する。課題 b については、以下の定式化が可能である。簡単のため、観測値  $o_i$  ( $i$  は最小部分領域) が Poisson 分布に従うと仮定する。なおその他の任意の分布についても、パラメータ推定値の算出部分を除くと、ほぼ同様の議論が成立する。さらに部分領域  $Z \subset G$  ( $G$  は全域) として、最小部分領域が任意の大きさに連結した領域 (「飛び地」のない連結領域: このような領域をウィンドウ  $Z$  と呼ぶ) を考える。このとき、ある  $Z$  における人口  $n_i$  あたりの  $o_i$  の発生率  $p_z$  を属性値としてみたとき、これが Hot spot か否かを同定する問題は、以下の仮説検定問題になる。

$$\begin{aligned} H_0 : p_z &= p_{z^c} \text{ for } \forall Z \subset G \\ H_1 : p_z &> p_{z^c} \text{ for } \forall Z \subset G \end{aligned} \quad (1)$$

ここで  $Z^c$  は部分領域  $Z$  の補集合である。また、

$$\begin{aligned} p_z &= o(Z)/n(Z) = \sum_{i \in Z} o_i / \sum_{i \in Z} n_i \\ p_{z^c} &= o(Z^c)/n(Z^c) = \sum_{i \in Z^c} o_i / \sum_{i \in Z^c} n_i \end{aligned} \quad (2)$$

は、Poisson 分布パラメータの最尤推定量である。

空間スキャン統計量は、 $H_0$  と  $H_1$  の 2 つのモデルの尤度比で定義される。Poisson 分布の各部分領域  $i$  の期待値  $e_i$  は、 $e_i = n_i \cdot o(G)/n(G)$  である。 $Z$  内外の期待観測度数は、 $e(Z) = \sum_{i \in Z} e_i$ 、 $e(Z^c) = \sum_{i \in Z^c} e_i$  となるのでこの尤度比は、

$$LR(Z) = \begin{cases} \left( \frac{o(Z)}{e(Z)} \right)^{o(Z)} \left( \frac{o(Z^c)}{e(Z^c)} \right)^{o(Z^c)} & o(Z) > e(Z) \\ 1 & o(Z) \leq e(Z) \end{cases} \quad (3)$$

で表される。以上より Hot spot 候補の  $Z$  は、ウィンドウ集合  $\tilde{Z}$  の中で最大の  $LR(Z)$  を持つ  $Z$ 、つまり、

$$\lambda_1 = \max_{Z \in \tilde{Z}} LR(Z) \quad (4)$$

となる  $Z_{(1)}$  として算出できる。同様の手順の適用により  $Z_{(1)}$  に次ぐ  $LR(Z)$  を持つ領域  $Z_{(2)}$  を、さらに  $Z_{(2)}$

1.ピークの検出	
P1	手持ちのデータ集合を、P1の実行が初めてのときは全て / P1の実行が2回目以降の場合は、以下のステップで既にピークに含まれる最小部分領域 <i>i</i> やこのステップで一度探索した最小部分領域 <i>i</i> を除く、と設定する。その集合から属性値が最大の最小部分領域 <i>i</i> を求める。なお手持ちのデータが空集合の場合は終了。
P2	<i>i</i> の近傍集合の中で、最大の属性値をとる近傍を <i>j</i> とする。 このとき、 <i>i</i> の属性値が <i>j</i> の属性値よりも大きければ、 <i>i</i> に <i>j</i> を結合してピークのエシエロンを形成する (形成順 <i>q</i> <sub>1</sub> , <i>q</i> <sub>2</sub> ,... と呼ぶ)。それ以外ならP1に戻る。
P3	処理中の <i>q</i> *の近傍集合の中で、最大の属性値をとる近傍を <i>k</i> とする (*はエシエロン番号を指す)。
P4	<i>q</i> *に <i>k</i> を結合した領域の近傍の中で最大の属性値をとる近傍を <i>m</i> とする。 このとき、 <i>k</i> の属性値が <i>m</i> の属性値以上ならば、 <i>q</i> *に <i>k</i> を結合した領域を改めて <i>q</i> *として、P3に戻る。 それ以外ならP1に戻る。
2.ファウンデーションの検出 / 1.を実行すると、どの <i>q</i> *にも所属しない最小部分領域 <i>i</i> が残る。なお1.で形成されたエシエロン数を <i>n</i> とする。	
F1	手持ちのデータ集合を、F1の実行が初めてのときは、すべての <i>q</i> *に属する最小部分領域以外の <i>i</i> の集合とする。2回目以降の時は、初期集合からこのステップで一度探索した最小部分領域 <i>i</i> を除く集合、と設定する。その集合から属性値が最大の最小部分領域 <i>r</i> を求める。なお手持ちのデータが空集合の場合は終了。 <i>r</i> が存在すれば、その <i>r</i> で新たにファウンデーションのエシエロン <i>q</i> *を形成する (形成順に <i>q</i> ( <i>n</i> +1), <i>q</i> ( <i>n</i> +2),... と呼ぶ)。 <i>r</i> が存在しなければ、アルゴリズムを終了する。
F2	F1で発見した <i>r</i> を近傍に持つ既に見つかったエシエロンは、 <i>q</i> *の子孫である。 <i>q</i> *の子孫とF1で発見した <i>r</i> を結合した領域を、 <i>q</i> *の一族とよぶ。
F3	処理中の <i>q</i> *の一族の近傍集合の中で、最大の属性値をとる近傍を <i>s</i> とする (*はエシエロン番号を指す)。 <i>s</i> が存在しなければ、F1に戻る。
F4	このとき、 <i>r</i> の属性値が <i>s</i> の属性値以上なら、 <i>q</i> *に <i>s</i> を結合した領域を改めて <i>q</i> *として、 <i>q</i> *の子孫と一族を更新して、F3に戻る。 それ以外ならF1に戻る。

に次ぐ  $LR(Z)$  を持つ領域を...のように計算を繰り返せば、 $K$  番目のウィンドウ  $Z_{(K)}$  まで順に探索できる。

課題 b には  $\lambda_1$  の分布が必要だが、それを解析的に求めることは困難である。そこで Kulldorff は、モンテカルロシミュレーションによって  $\lambda_1$  の分布を求める方法を提案した。具体的には、まず人口  $n_i$  を固定して観測値  $o_i$  をランダムに並べ替えた疑似データを作成し、元データと同じウィンドウ集合  $\tilde{Z}$  の中で尤度比が最大となるウィンドウの  $\lambda_1^1$  を算出する。Hot spot は、以下の手順によって同定する。ウィンドウの探索によって同様の作業を  $N$  回繰り返して得られる尤度比の集合

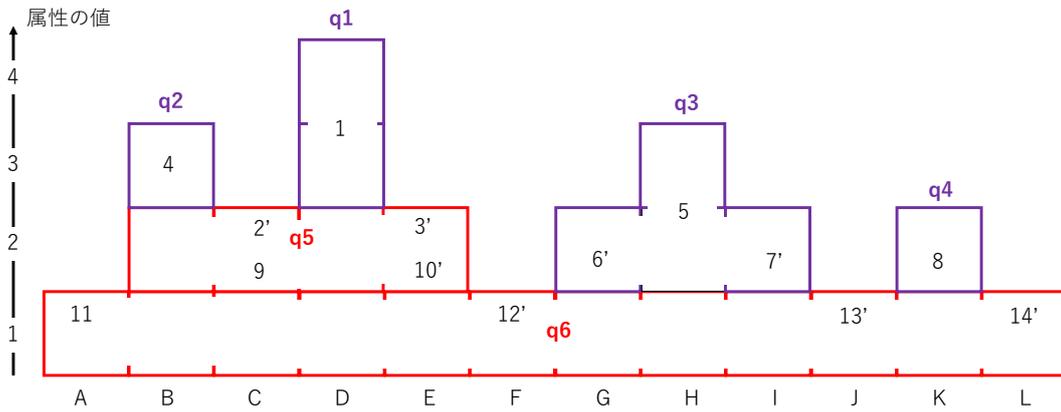


図 1 1次元エシェロン探索の様子

$\Lambda = (\lambda_1^1, \dots, \lambda_1^n)$  を降順に並べたとき、最初に得た  $\lambda_1$  が挿入される順位  $R$  を求める。このとき、ホットスポット候補  $Z_{(1)}$  の疑似  $p$  値は、

$$p = \frac{R}{1+N} \quad (5)$$

として算出できる。ここで  $N$  は繰り返し回数を表す。

疑似  $p$  値を分かりやすく表現するため、アルゴリズムを実装する際には、 $N=999, 9999, 99999$  などが用いられる。さらに Kuldorff は、 $Z_{(2)} \sim Z_{(K)}$  の疑似  $p$  値は、 $Z_{(1)}$  の疑似  $p$  値を求めたときと同じ  $\Lambda$  を参照することによって、検定の多重性が生じない、と指摘している。

空間スキャン検定は、一般に Hot spot となるウィンドウ  $Z$  の発見、ならびにウィンドウ集合  $\tilde{Z}$  の設定が難しいという問題がある。エシェロン解析は、そのようなウィンドウを効率的に探索するアルゴリズムであり、課題 a の実行可能解の一つである。エシェロン解析の概要を表 1 に示すとともに、1次元のエシェロン探索を図 1 にまとめる。

図 1 は、A~L の 12 の最小部分領域からなる領域 G において、属性値が 1 から 4 の値を持つことを示している。同図の線分上で隣り合う部分領域は、相互に近傍と考える。表 1 に示すアルゴリズム P1~P5 を適用すると、「近傍よりも高い属性値を示す隣接する最小部分領域の集合」であるピークエシェロンが得られる。図中の数値は探索順序を示しており、「'」が付されていない探索順序で新しくピークエシェロンが形成される。その結果、同図では、q1~q4 が得られる。なお属性値が同値のときは、左側から探索することとした。また、P1~P5 の探索では、8 番目 (K) の次は、9 番目に A、10 番目に F、11 番目に J、12 番目に L を探索するが、いずれも P4 の条件を満たさないため、ピークエシェロンに編入されない。ピークエシェロンを「支える」下部は、エシェロン解析で

はファウンデーション (エシェロン) と呼ばれる。これらを検出する F1~F5 では、q1~q4 に属する最小部分領域は探索しないので、A, C, E, F, J, L の 6 領域が探索対象になる。F1 より C が最初に探索され (探索順序 9)、ファウンデーションを形成する (q5)、E (探索順序 10) は q5 に編入される。q5 の「上に載っている」エシェロン q2 と q1 は q5 の子孫、q2 と q1 にとって q5 は親、そして q2, q5, q1 は一族と呼ばれる。A (探索順序 11) は左側探索のためその次に探索され、F1 よりファウンデーションを形成し (q6)、残る F, J, L は q6 に編入される。

なお同図では、ファウンデーションの一族に子孫のエシェロンを含むことを明確にするため、属性値が同値の場合は、一族全体を含むようにファウンデーションを表記する (q5, q6)。他方で最小部分領域でエシェロンを表記すると、q5=(C, D)、q6=(A, F, L) となることに留意されたい。さて、以上のようにエシェロン間の親子関係を定義することによって、q6 を全体の親 (=根)、ピークエシェロン (q1, q2, q3, q4) を末子孫 (=葉) とする、階層図 (デンドログラム/樹形図) が描ける。以上の手順から明らかなように、P1~P5、および F1~F5 の適用回数は、全域の属性値が同値の場合でも高々  $2M$  ( $M$  は最小領域数) 回に過ぎず、近傍リストさえ構築すれば 2次元や 3次元、あるいはグラフ構造への適用も容易である。またピークエシェロンの近傍となるファウンデーションの部分領域が、近傍となるピークエシェロンに均等に所属することを認めれば、ファウンデーションを各ピークエシェロンに所属させ、相互排反かつ全領域を網羅するような「圏」を算出することもできる (A~C, C~F, F~J および, J~L)。このようにエシェロン解析は、最小部分領域を効率的、かつシステムティックに分類できるという利点がある。

石岡・栗原が提案したエシェロンスキャン法は、空間スキャン検定の  $Z$  ならびに  $\tilde{Z}$  の設定に、エシェロン解

析を組み込んだ方法であり、以下の手順で実行する。

- 1) 解析対象領域  $G$  のエシェロンデンドログラムを作成する。
- 2) エシェロンデンドログラムの上位階層（ここでは属性値が最大のエシェロン）から、その階層に含まれる領域を上から順に  $Z$  に加えて  $\tilde{Z}$  を構成する。なお効率化を図る場合、Kulldorff は一つの  $Z$  内の人口が  $G$  の人口の 50% を超えるまで、 $\tilde{Z}$  に加え続けることを推奨している。
- 3)  $\lambda_1 = \max_{Z \in \tilde{Z}} LR(Z)$  となる  $Z_{(1)}$  を第一 Hot spot 候補とする。
- 4) モンテカルロ法によって Hot spot の有意性を評価する。

以上の手順から得られる  $Z$  は、必ず飛び地の無い、相互に連結した領域群となる。

### 3. 都心地区来訪者分析の方法

本研究では、モバイル空間統計によって得られる 500m メッシュ単位の人出データと都市施設の空間分布データを使用する。実際の都心域にエシェロンスキャン検定を適用する際に、 $o_i$  に人流データを当てることは自明である。しかし基準化変数  $n_i$  の設定には、工夫が必要である。候補データは人口や従業員数のほか、様々な都市施設数、従業員数、事業所数など複数のデータが考えられる。ただし集客ポテンシャルのある都市施設は複数考えられるので、予備分析として基準化変数候補に関する主成分分析を適用して、抽出した主成分と人流の相関が最も高くなる重みで線形結合した変数で基準化するか、または回帰主成分分析などを工夫する必要がある。

別のアプローチとして、人口などを用いて人流と都市施設それぞれに関して変数の基準化を行って、いずれかのエシェロン解析から得られた領域を、共通して適用する方法が考えられる。通常は異なる変数から得られるピークエシェロンやファウンデーションは一致しないが、都市施設が存在する地点における人流のミスマッチを解明する手法を開発するという本題の趣旨を踏まえると、

都市施設のエシェロンを用いるべきと思われる。すなわち都市施設の解析から得られるエシェロン集合に関して、人口で基準化した都市施設の Hot spot 対数尤度と、人口で基準化した人流の Hot spot 対数尤度の散布図を描いて、両者の比例関係からの逸脱度を検証する手順が考えられる。この際、候補となる都市施設ごとに Hot spot 対数尤度を算出して比較することになる。

いずれの方法を用いる場合も、基準化後の変数の分布に対して、適切な確率分布を想定する必要がある。この点は、分布の適合度を事前に検定する必要がある。

### 4. 分析結果

発表時に報告する。

#### 参考文献

- 1) 栗原考次, 石岡文生: エシェロン解析 (統計学 One point), 共立出版, 2021.
- 2) 小田牧子, 石岡文生, 正木隆, 栗原考次: エシェロン階層構造を利用した森林の分割, データ分析の理論と応用, Vol.2, No.1, pp.17-31, 2012.
- 3) 堤盛人, 瀬谷創: 応用空間統計学の2つの潮流: 空間統計学と空間計量経済学, 統計数理, Vol.60, No.1, pp.3-25, 2012.
- 4) 石岡文生, 栗原考次: Echelon 解析に基づくスキャン法によるホットスポット検出について, 統計数理, Vol.60, No.1, pp.93-108, 2012.
- 5) 楯西将司, 石岡文生, 栗原考次: 空間疫学におけるリスク評価とその可視化のためのソフトウェア, 計算機統計学, Vol.33, No.2, pp.91-105, 2020.
- 6) Tango, T. and Thakahashi, k.: A flexible spatial scan statistic with a restricted likelihood ratio for detecting disease clusters, Statistics in medicine, vol.31, pp.4207-4218, 2012. DOI: 10.1002/sim.5478.
- 7) Kulldorf, M.: A spatial scan statistics, Communications in statistics: Theory and Methods, Vol.26, No.6, pp.1481-1496, 1997.

(Received??, ????)  
(Accepted ??,???)

Analysis of intercity passenger arrivals to city center areas based on echelon analysis

Makoto TSUKAI, and Ryo ICHIMANDA

Abstract