

交通事故発生地点の経度・緯度情報を用いた事故分析

山田 晴利¹・下村 静喜²・田中 祥夫³・柴崎 亮介⁴

¹正会員 博士（工学） 交通事故総合分析センター 常務理事

（〒101-0064 東京都千代田区猿楽町2-7-8）

E-mail: yamada@itarda.or.jp

²非会員 交通事故総合分析センター 研究第2課 （〒101-0064 東京都千代田区猿楽町2-7-8）

E-mail: s_shimomura@itarda.or.jp

³非会員 交通事故総合分析センター 研究部 次長 （〒101-0064 東京都千代田区猿楽町2-7-8）

E-mail: y_tanaka@itarda.or.jp

⁴正会員 工学博士 東京大学教授 空間情報科学研究センター （〒277-8568 千葉県柏市柏の葉5-1-5）

E-mail: shiba@ccsis.u-tokyo.ac.jp

2012年から交通事故発生場所の経度・緯度が計測されるようになり、交通事故の地理空間的な分析を行うことが可能になった。本研究では2013年の人身事故の中から幹線的な一般道路で発生した事故を抽出し、道路リンク毎の件数を求めて多発箇所を特定し、東京都市圏と京阪神都市圏を対象に死亡事故、重傷事故、多発箇所の分布パターンの中に差があるかどうかを検証した。このいずれの組合せにおいてもパターンの中に有意な差があることが判明した。さらに、1 kmメッシュ単位に集計した事故件数を用いて都道府県別に事故のクラスタリング分析を行ったところ、都市部では事故が多発しているクラスターが存在する一方、非都市部ではクラスターは存在しないことが認められた。

Key Words: traffic accidents, longitude and latitude, spatial pattern, clustering

1. 目的

2012年からわが国の人身交通事故の発生場所の経度・緯度が計測され、事故原票にそのデータが付与されるようになった。本研究では、事故多発箇所（ホットスポット）と重大事故（死亡事故と重傷事故）の分布パターンに違いがあるかどうかを分析すること、および交通事故をメッシュ単位に集計してメッシュのクラスタリングを見出すことを目的として分析を行った。

諸外国ではこれまでこうした分析がいくつか行われているが（たとえば、Kuo *et al.*¹⁾）、わが国では事故発生場所の位置情報が把握されていなかったため、行われていない。

なお、計測された事故発生場所の経度・緯度の正確さには都道府県によってばらつきがあるが、多くの都道府県では概ね正確であることがわかっている²⁾。特に、本研究で分析対象とした都道府県では正確な値が与えられており、分析に用いても問題はないと判断した。

2. 利用データ

2013年に発生した人身交通事故（死亡、重傷及び軽傷事故）の中から幹線的な一般道路（一般国道、主要地方道、および一般都道府県道）で発生した事故を抽出し、後述のITARDA区間毎に件数を集計した。物損事故については、発生場所の経度・緯度が計測されていないため分析対象としなかった。また、高速道路、自動車専用道路、市町村道、農林道、広場・駐車場等での事故は除外されている。

ITARDA区間は道路管理者が交通事故の集計・分析に利用している区間であり、単路と交差点から構成されている。単路の平均長は約300 mであり、全国で単路区間が約60万、交差点が約24.8万箇所存在する。

これらITARDA区間別に人身事故件数を集計し、事故件数が4以上の区間を多発箇所とした。多発箇所の数は全国で約1万7千であった。以下の分析では、多発箇所の場所を表す代表点として、交差点についてはデジタル道路地図の交差点の代表点を、単路については始点と終点

の中点をそれぞれ用いた。

また、人身事故をメッシュ単位に集計する際には1 km メッシュ（第3次地域メッシュ）を用いた。

3. 重大事故と多発箇所の分布パターンの比較

東京都市圏と京阪神都市圏を対象に、死亡事故、重傷事故および多発箇所の分布パターンに差があるかどうかを検証した。東京都市圏では三鷹市を中心とした半径30 kmのエリアを、京阪神都市圏では守口市を中心とした半径50 kmのエリアをそれぞれ分析対象とした。円の中心位置と半径は、海域ができるだけ少なくなる一方で、都市地域をできるだけ多く含むように試行錯誤的に設定した。各都市圏における死亡事故、重傷事故および多発箇所の分布とカーネル密度推定の結果を図-1と図-2に示した。

東京都市圏では、東京都心部で死亡事故の密度が高く、分布は南北方向に広がっている。一方、重傷事故は東京での密度が低く、埼玉県南部で密度が高くなっている。多発箇所は東京都心の密度が高く、分布は東西方向及び南部方向に広がっている。重傷事故の分布が埼玉県南部に偏っているのは、東京都の重傷事故の件数が周辺の県と比べて著しく低いためである。しかし、東京都で重傷事故件数が少ない原因は不明である。

京阪神都市圏では、死亡事故、重傷事故及び多発箇所いずれの場合でも大阪都心部の密度が高く、京都都心部にもピークが見られる。東京都市圏で見られたような重傷事故の分布の偏りは見られない。

東京都市圏と京阪神都市圏それぞれにおいて、死亡事故、重傷事故および多発箇所を二つずつ組み合わせ、マーク付き点過程の分析手法を使って分布パターンに差が

あるかどうかを検定した。

空間分布パターンに差があるかどうかを検定するのに「クロスL関数」を使った³⁾。クロスL関数は次式の「クロスK関数」から計算される。クロスK関数はマークが*i*の点からある一定の距離*r*以内にあるマークが*j*の点の数をもとにして計算される。すなわち、

$$\text{Cross } K_{ij}(r) = \lambda_j E[K_{ij}(r)]$$

ここに λ_j はマークが*j*の点の平均的な密度、 $K_{ij}(r)$ はマークが*i*の点から一定の距離*r*以内にあるマークが*j*の点の数で、 $E[\cdot]$ は期待値を表す。クロスL関数は次式で定義される。

$$\text{Cross } L_{ij}(r) = \sqrt{\text{Cross } K_{ij}(r)/\pi}$$

マーク*i*の点とマーク*j*の点のパターンが独立ならば、クロスL関数の理論的な値は*r*と与えられる。すなわち、距離*r*の線形関数となり、クロスK関数より安定しているといわれる。

従来は、点パターンの分析では、シミュレーションを行って包絡線を計算し、パターンに差があるかどうかを判定するという手続きがとられていた。一例として、東京都市圏の死亡事故と重傷事故を対象に包絡線を計算した結果を図-3に示した。この図中の赤い点線は理論分布（CSR: Complete Spatial Randomness）から求めたクロスL関数、灰色の領域の上側の線がシミュレーションで求めたクロスL関数の最大値、下側の線が最小値である。観測データから求めたクロスL関数はシミュレーションで求めた最大値より大きく、したがって死亡事故と重傷事

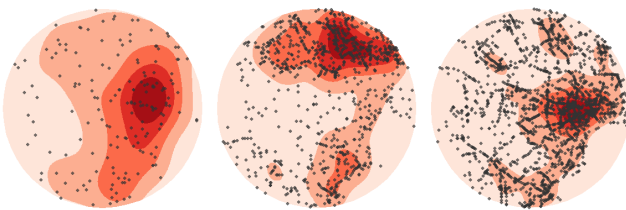


図-1 東京都市圏における死亡事故（左）、重傷事故（中）および多発箇所（右）のカーネル密度推定結果

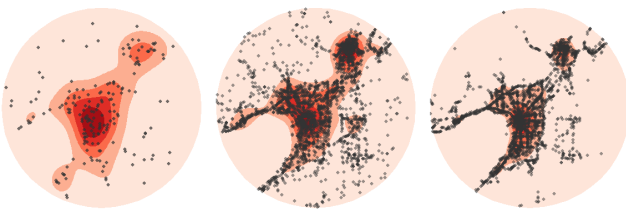


図-2 東京都市圏における死亡事故（左）、重傷事故（中）および多発箇所（右）のカーネル密度推定結果

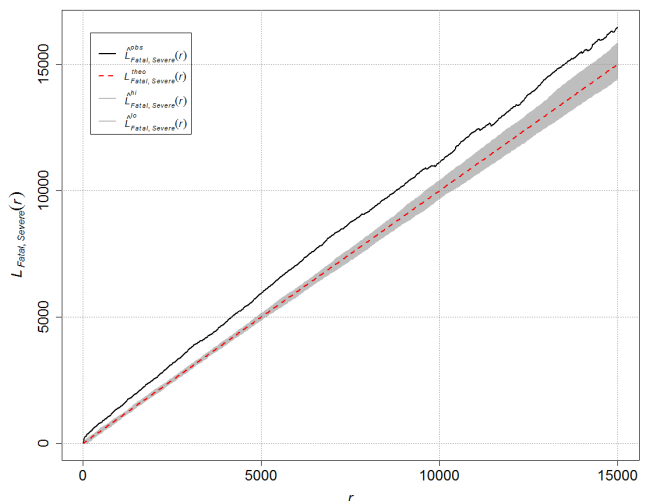


図-3 東京都市圏における死亡事故と重傷事故のクロスL関数と包絡線

故の分布パターンの間には「有意な差がある」と結論づけるのが従来のやり方である。しかし、この方法は多重比較を行っているので、検定の方法としては望ましくない (Brunsdon *et al.*³⁾, Baddeley *et al.*⁴⁾)。

包絡線を使った検定に変わる方法として考案されたのが、DCLF検定とMAD検定であり、一つの検定統計量を用いる。すなわち、DCLF検定では検定量として

$$u_i = \int_0^R [H(r) - H_{theo}(r)]^2 dr \\ \approx \sum_{t_k=t_{min}}^{t_{max}} [\hat{H}_i(t_k) - \bar{H}_i(t_k)] \delta t$$

を使う。ここに、 t_k は距離の分点、 t_{min} と t_{max} は距離の最小値と最大値、 $\hat{H}_i(t_k)$ は観測された空間データ*i*の検定統計量（ここではL関数）の値、 $\bar{H}_i(t_k)$ は*i*を除いて計算した期待値、 $\delta t_k = t_{k+1} - t_k$ は区分された距離の幅である。 $\bar{H}_i(t_k)$ は次式で計算される。

$$\bar{H}_i(t) = \frac{1}{s-1} \left\{ \sum_{j=1, j \neq i}^s \hat{H}_j(t) \right\}$$

s はシミュレーションの回数を表し、番号1は観測されたデータに対応しており、シミュレーションには2から*s*までの番号が振られているものとする。大数の弱法則によれば、 $\lim_{s \rightarrow \infty} \bar{H}_i(t) = H(t)$ がなりたつ ($H(t)$ は検定統計量の理論値)。したがって、 u_i は観測された点パターンと理論的な結果の差の二乗を関心のある距離の範囲で積分した値となっている。

DCLF検定の対象となる帰無仮説 H_0 は、距離 $t \in [t_{min}, t_{max}]$ において計算された観測データの統計量にもとづく u_i の値が仮定された空間過程のモデルからランダムに生成された値と有意に異なっていない、というものである。

MAD検定量は次式で計算できる

$$MAD = \max_d |\hat{H}(d) - H_{CSR}(d)|$$

ここに、 $H_{CSR}(d)$ は点過程がCSRであると仮定したときの関数 H の値である。

Baddeley *et al.*⁴⁾は、シミュレーションを使ってDCLF検定とMAD検定の比較を行い、DCLF検定の方がよいと結論づけている。

以上の記述を踏まえて、東京都市圏と京阪神都市圏を対象に事故多発箇所、死亡事故発生箇所および重傷事故

発生箇所の分布パターン間に統計的に有意な差があるかどうかを検定した。シミュレーション回数は499回とした。

この結果によると、東京都市圏と京阪神都市圏いずれにおいても、事故多発箇所と死亡事故発生箇所、重傷事故発生箇所の間には有意な差があるという結論であった ($p < 0.01$)。この結果は、死亡事故、重傷事故は必ずしも事故多発箇所が発生していないこと、そして死亡事故と重傷事故も必ずしも同じ場所で発生しているわけではないを意味している。したがって、死亡事故、重傷事故を減少させるためには、事故多発箇所以外の場所での対策も重要であるという結論が導かれる。

4. 交通事故のクラスタリング分析

1 kmメッシュ単位に集計した事故件数を持ちいて、都府県別（埼玉県、東京都、神奈川県、京都府、大阪府、奈良県、兵庫県）に事故のクラスタリングの分析を行った。平日と休日の事故を分けて分析したが、メッシュ単位で集計した事故パターンには平日と休日の間で大きな差異が見られなかったため、以下では平日の事故についての結果のみを示す。

まずデジタル道路地図を用いてそれぞれのメッシュに存在する道路延長を求め、道路延長が0のメッシュは分析対象から除外した。これは道路延長が0のメッシュでは特殊な交通事故以外は発生しないと予想されるからである。地方部では道路延長が0のメッシュの数が多いように、事故件数が0のメッシュの数も多い。

クラスタリングの分析ではあるメッシュの近傍にあるメッシュを定義する必要があるため、ここでは重心間の距離が2 km以下のメッシュを近傍にあるものと定義した。さらに次式で定義されるOrdのG統計量の値をメッシュ毎に計算し、分析に用いた³⁾。

$$G_i = \frac{\sum_{j=1}^n z_j v_{ij}(d)}{\sum_{j=1}^n z_j}$$

ここに、 $v_{ij}(d)$ はポリゴン*i*と*j*のセントロイド間の距離が*d*未満のとき値1を、そうでなければ値0をとる関数であり、 z_i は関心のある観測量（ここでは事故件数）である。上の式からわかるように、G統計量はあるメッシュとその近傍のメッシュの事故件数が全体の事故件数に占める割合を表す。ただし、観測量の値は正である必要がある。

従来、G統計量を用いて統計的な検定を行うために、次の手順がとられていた。

- (1) それぞれのポリゴンごとにG統計量の値 G_i を計算

する。

- (2) 理論的に求められる G_i の期待値と分散を用いて、 p 値を計算する。
- (3) p 値をもとにして有意性を判断する。

しかしながら、この手順では、ポリゴンごとの G_i が正規分布すると仮定されていることに加えて、多重比較が行われているという問題が存在する。正規性の分布は必ずしも成立しないことから、ここでは上記の手順を採用せず、「各ポリゴンの事故件数を任意の順番に並べ替えて得られる仮想的なパターンはどれも等しい確率で起こる」という仮説にもとづいて、シミュレーションによって G 統計量を計算し、 p 値を求めた（多重検定になるのでFDR法で補正した）。神奈川県と京都府の計算結果を図4と5に示した。どの都府県でも都市部と幹線道路沿いに事故件数の多いクラスターが集中しており、非都市部には存在しないことがわかる。なお、Kuo *et al.*¹⁾は、交通事故が多発しているメッシュと犯罪が多発しているメッシュは重なることが多いので、こうした地域で重点的に取り締まりを行えば、事故と犯罪双方の抑止に効果があると述べている。こうした分析については今後の課題としたい。

5. おわりに

本報告では、交通事故の発生場所の経度・緯度情報を使って事故の分布パターンとクラスタリングを分析した。これらの分析には統計ソフトRを用いた。

参考文献

- 1) Kuo, P-F., Zeng, X. and Lord, D.: Guidelines for choosing hot-spot analysis tools based on data characteristics, network restrictions, and time distributions, Presented at the 91st Annual Meeting of the TRB, November, 2011.
- 2) 山田晴利, Teerayut Horanont, 田中祥夫, 柴崎亮介: 交通事故発生場所の経度・緯度の精度検証と事故分析システムの開発, 土木計画学研究・講演集 vol. 49, 9 pages, 2014.
- 3) Brunson, C. and Comber, L.: *An introduction to R for spatial analysis & mapping*, Sage, 2015.
- 4) Baddeley, A., Diggle, P. J., Hardegen, A. H., Lawrence, T. Milne, R. K. and Nair G.: On tests of spatial pattern based on simulation envelopes, *Ecological Monographs*, 84(3), 477-489, The Ecological Society of America, 2014.

(2016. 4.22 受付)

A SPATIAL ANALYSIS OF TRAFFIC ACCIDENTS USING THEIR LONGITUDE AND LATITUDE OF OCCURRENCE

Harutoshi YAMADA, Shizuki SHIMOMURA, Yoshio TANAKA
and Ryosuke SHIBASAKI

Spatial pattern of traffic accidents has been analyzed using their longitude and latitude of occurrence. This locational information has become available since 2012 and is very useful for spatial analysis. In this paper, two kinds of analyses, one on the differences among spatial patterns of fatal, severely injured accidents and hot spots and the other one on the clustering of traffic accidents have been made.

The spatial patterns of fatal, severely injured accidents and hot spots have been proved to be non-identical. This indicates that traffic safety improvement should not be concentrated on hot spots. In addition, it has been found that clustering of accidents exists only urban areas and that they does not exist on rural areas.

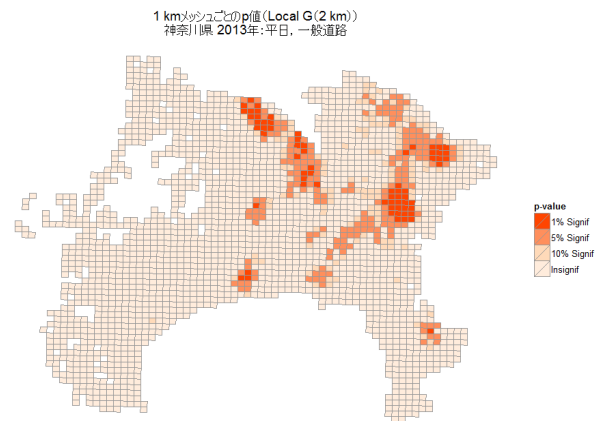


図4 神奈川県の交通事故のクラスタリング分析結果
(2013年, 一般道路)

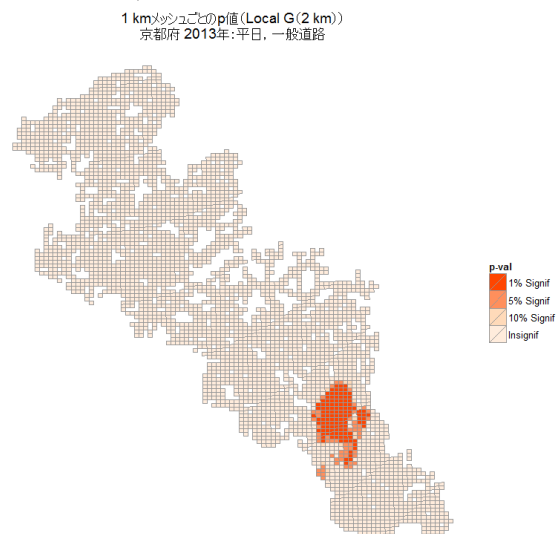


図5 京都府の交通事故のクラスタリング分析結果
(2013年, 一般道路)