

高速道路リンクの交通事故発生モデルの開発

山田 晴利¹・Quanjin Chen²・田中 祥夫³・柴崎 亮介⁴

¹正会員 博士（工学） 特任教授 東京大学空間情報科学研究センター（〒153-8505 東京都目黒区駒場 4-6-1）

前常務理事 公益財団法人交通事故総合分析センター

E-mail: yamada.hal@csis.u-tokyo.ac.jp

²非会員 東京大学大学院（〒277-8568 千葉県柏市柏の葉 5-1-5）

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

⁴正会員 工学博士 教授 東京大学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）

交通事故データでは件数が 0 となるリンクやエリアが多数発生するため、モデル化が難しいことが知られている。この問題を解決するために、従来負の二項分布モデルやゼロ強調モデルが使われてきたが、必ずしも満足のいく結果が得られていない。本研究では、高速道路のインターチェンジ間リンクにおいて発生した事故件数を分析対象としてとり上げ、隣接するリンクの間には空間的な相関関係が存在すると仮定し、条件付き自己回帰モデル（CAR）に従うランダム効果項を組み込んだポアソン回帰モデルを MCMC によって推定した。CAR としてはこれまで広く使われている intrinsic CAR ではなく、それを一般化した Leroux 他モデルを用いた。その結果、ゼロ強調モデル等と比べ優れたモデルを構築することができた。

Key Words: traffic accidents, expressways, Poisson regression models, conditional autoregressive process

1. はじめに

交通事故の発生要因を探り、さらには将来の事故件数を予測する上で事故発生モデルの果たす役割は大きい。しかしながら、わが国ではこれまで事故の発生場所の情報が整備されていなかったため市区町村以下のエリアあるいは道路のリンク単位でのモデル構築が難しかった。2012 年からは都道府県警察によって事故発生場所の経度・緯度情報が取得されるようになった¹⁾ので、ここではこのデータを用いて高速道路のリンク単位の事故発生モデルの構築を試みた結果を報告する。

ただし交通事故の分析に当たっては、事故発生場所の特定に加えて以下のような問題が存在する。

一つは事故件数 0 のメッシュあるいはリンクが大多数を占めることであり、もう一つは重要な事故発生要因である自動車や歩行者の交通量データが十分に整備されていないことである。

最初の問題の解決のために、複数年の件数を合計して分析に用いる、あるいはゼロ強調モデル等を使ってモデルを構築するといった方策がとられることが多い。しかし複数年の件数を合計する方法には、次のような問題点がある。すなわち、重大事故（死亡事故や重傷事故）が発生した箇所では、発生後速やかに事故対策が行われることが多く、事故発生が抑止される。したがって、こうした箇所の事故情報を使い続けることは、爾後の分析結果を誤った方向に導く可能性がある。

また、事故件数が 0 のデータが多いと「過分散」

（overdispersion）の問題が発生する^{2),3)}。過分散の問題は次のように記述できる。交通事故はまれに発生する事象なので、そのモデル化ではポアソン分布が使われることが多い。しかし実際の事故データでは分散の値が期待値より大きく、期待値と分散の値が等しいポアソン分布では満足のいく結果が得られない。

このため、ポアソン分布ではなく、負の二項分布あるいはゼロ強調モデル等が使われることがある^{4),5),6)}。負の二項分布、ゼロ強調モデルでは、分散の値は期待値より大きくなるので、過分散の問題を解消できる。

しかしながらこれらのモデルを使ったとしても、空間的な自己相関を表現できないという問題が残る。空間的な自己相関とは、事故の多発している（あるいは事故件数の少ない）メッシュあるいはリンクが空間的に集中しクラスターを形成していることをいう⁷⁾。前述のポアソンモデル、ゼロ強調モデルではデータは互いに独立であると仮定されており、こうした空間的な自己相関を考慮することができない。

実際、本分析で用いる事故データについてモランの I 等の指標を使って空間的なクラスターが存在するかどうかを検定したところ、クラスターが存在しないという帰無仮説は棄却されることがわかった。

空間的な自己相関をモデルにくみ込むために、回帰モデルの中にランダム効果項を取り入れ、ランダム効果項が空間的な自己相関をもつ事前分布にしたがうという前提の下でモデルを構築するという方法がとられる^{3),8)}。したがって、モデルを推定するにはマルコフ連

鎖モンテカルロ法 (McMC) を使うことになる。

ランダム効果項に与える事前分布としては、従来 *intrinsic CAR* が広く使われてきた⁹⁾。しかし、Lee¹⁰⁾ がシミュレーションを使って分析した結果によれば、*intrinsic CAR* は強い空間的自己相関がある場合には適しているが、空間的自己相関が弱い場合には良いモデルとはいえないことがわかっている。Lee¹⁰⁾ は、自己相関が弱い場合でも強い場合でも適用できる *Leroux et al.* のモデル¹¹⁾ を推奨しているので、以下の分析でも *Leroux et al.* のモデルを使うことにする。*Leroux et al.* のモデルは特別な場合として *intrinsic CAR* を含んでおり、*intrinsic CAR* よりも一般的なモデルとなっている。

次に、交通事故分析における二つ目の問題としてとり上げた交通量データが十分に整備されていない点について述べる。

事故の分析では、自動車、歩行者、さらには自転車、二輪車等の交通量が要因として重要な役割を果たす。しかし、実際にはこうした交通量データは十分に整備されておらず、我が国で利用可能なデータとしては道路交通センサスの調査結果程度しかないのが現状である。しかも、道路交通センサスでは5年に一度、平均的な交通量が出現すると思われる一日の交通量が計測されているにすぎない。さらに、交通量が観測されているのは幹線道路に限定されており、市町村道等の交通量は計測されていないという問題がある。

こうした制約があるため、プローブデータや携帯・スマホの履歴データを利用して交通量を推計する試みが行われている^{12),13),14)}。しかし、精度の高い推計データはまだ存在しないのが実状なので、本研究では道路交通センサスのデータを利用可能な高速道路での交通事故について分析を行うこととした。

本報告の構成は次のとおりである。第2節で分析に用いる交通事故データの概要と基本的な統計量を提示する。第3節では、*Leroux et al.* のモデルについて述べ、*intrinsic CAR* との違いを示す。第4節で *Leroux et al.* によるモデルを推定した結果を述べるとともに、他のモデルとの比較を行う。最後に第5節で本研究の成果をまとめ、考察を加える。

2. 分析に用いるデータと基本統計

本研究では、2014年に高速道路で発生した人身事故をリンクごとに集計した事故データを分析の対象とした。リンクはインターチェンジとインターチェンジを連結する区間ごとに一つ設定した。事故の重度(死亡、重傷および軽傷)、事故類型は区別していない。ただし、平日に発生した事故のみを分析の対象とし、休日に発生した事故は対象外とした。

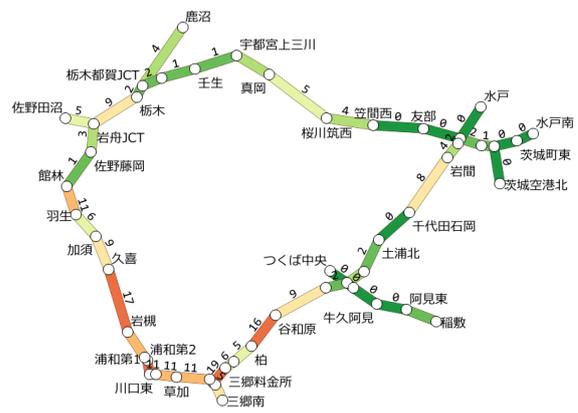


図-1 分析対象とした高速道路ネットワーク：○はインターチェンジを表し、リンクの横の数字は2014年平日の人身事故件数である

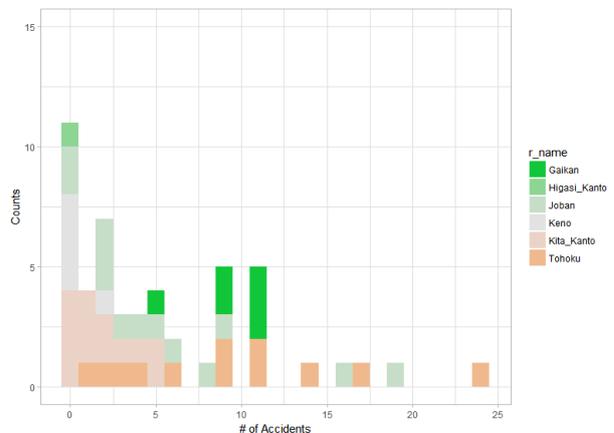


図-2 分析対象リンクの事故件数のヒストグラム：路線別に色分けしてある

分析対象としたリンクは、外環道、東北道、北関東道、東関東道および常磐道からなる50のリンクである。図-1に分析対象のネットワークと事故件数を示した。

各リンクについて、リンク延長、車線数、規制最高速度、車道幅員、中央帯幅員、沿道状況、2010年の道路交通センサスによる平日24時間交通量を「平成22年度道路交通センサスDVD-ROM¹⁵⁾」から抽出し、説明変数として用いることとした。事故のデータ取得年に合わせて2015年のセンサス交通量を使いたかったのであるが、研究時点ではこのデータが未公開だったため、2010年のセンサス交通量で代替した。

分析対象としたリンクの事故件数についてヒストグラムを描いた結果を図-2に示した。この図では事故件数が路線別に色分けされており、圏央道、北関東道では件数が少ない(5件以下)ことが見てとれる。これに比べると東北道、常磐道では事故の件数が多い。また、事故件数がゼロのリンクの数が最も多くなっているが、件数ゼロのリンク数が極端に多いわけではない。なお、

表-1 リンクの事故件数の空間的自己相関

統計量	値	p 値
モランの I	0.563	0.001
Geary の C	0.391	0.001
Getis-Ord の G	0.073	6.8e-5

事故件数の平均値は 5.38, 分散の値は 32.3 となっており, 分散の値が平均値の約 6 倍となっている。

さらに, リンクの空間的自己相関を分析するために, 次のようにして隣接行列 W を定義した. すなわち, リンク i, j が互いに連結されているときには $w_{ij} = 1$, そうでないときは $w_{ij} = 0$ とした. (なお, $w_{ii} = 0$ である.)

空間的自己相関を表す指標には, モランの I, Geary の C, Getis-Ord の G がある^{7,8)}. 隣接行列 W と各リンクの事故件数 y_i を使ってこれらの値を計算した結果を表-1 に示した. この表には計算された統計量の値について, データのランダムな並べ替えによるシミュレーションで求めた p 値も掲げている。

表-1 に示した三つの統計量の値は, いずれも事故件数の高い値を持つリンクが空間的にクラスタリングしているという結果を示している. したがって, 空間的自己相関を考慮に入れた分析を行う必要がある。

3. 空間的自己相関をもつモデル

各リンクの交通事故の件数 Y_i を以下のポアソン回帰モデルでモデル化する。

$$Y_i \sim \text{Poisson}(\mu_i), \log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} + O_i + \phi_i. \quad (1)$$

ここに, μ_i はポアソン分布の期待値, \mathbf{x}_i は説明変数からなるベクトル, $\boldsymbol{\beta}$ は回帰係数のベクトル, O_i はオフセット項, ϕ_i はランダム効果項である. ランダム効果項 ϕ_i については, フルコンディショナルな事前分布が Leroux *et al.*¹¹⁾ によって与えられた次式に従うものと仮定する。

$$\phi_i | \phi_{-i}, W, \tau^2, \rho \sim \mathcal{N} \left(\frac{\rho \sum_{j=1}^n w_{ij} \phi_j}{h_i}, \frac{\tau^2}{h_i} \right). \quad (2)$$

ここに, $h_i = \rho \sum_{j=1}^n w_{ij} + 1 - \rho$ で, \mathcal{N} は正規分布を表し, τ^2 は正規分布の分散パラメーター, ρ はモデルパラメーターで 0 と 1 の間の値をとる. $\rho = 1$ の場合これまで広く使われてきた intrinsic CAR モデルになる. また, ϕ_{-i} は i 以外のリンク集合で, n はリンク総数を表す。

Brook の補題⁸⁾ を使って式 (2) で与えられるフルコンディショナルな分布から $\boldsymbol{\phi}$ の同時分布を求めると, 次式のようになる。

$$\boldsymbol{\phi} \sim \mathcal{N} \left(\mathbf{0}, \tau^2 [\text{diag}(W\mathbf{1}) - W]^{-1} \right). \quad (3)$$

ここに, $\mathbf{1}$ は n 個の 1 からなる列ベクトル, $\text{diag}()$ は括弧内の要素を対角要素にもつ対角行列を意味する。

式 (2) で $\rho = 0$ とおくと ϕ_i のフルコンディショナルな分布の期待値が 0 となり, ランダム効果項の間の相関が弱い状態を表現できることがわかる. 一方, $\rho = 1$ とおけば強い相関関係を表現することができ, Leroux *et al.* のモデルは弱い相関関係から強い相関関係までを表現できることがわかる。

モデルの推定には統計ソフト R¹⁶⁾ を利用し, McMC の計算には Stan¹⁷⁾ を用いた. なお, リンクの数が増えるのにもとない隣接行列 W の次元も大きくなっていき, 式 (3) に示した同時分布を使って事前分布 (の対数) の値を計算しようとする, 計算に要する時間も長くなる. この問題を解決する方法が Joseph¹⁸⁾ によって開発されているので, ここでもその方法を利用した。

表-2 事故件数と説明変数の間の相関係数

変数	事故件数	延長	車線数	最高速度	車道幅員	中央幅員
延長	-0.004					
車線数	0.391	0.137				
最高速度	-0.185	0.383	0.547			
車道幅員	0.368	0.174	0.991	0.636		
中央幅員	0.157	0.192	0.744	0.821	0.815	
交通量	0.681	-0.059	0.754	0.063	0.718	0.411

4. モデルの推定結果と考察

まず, 説明変数と目的変数の間の相関係数を求めた結果を示す (表-2). この表中の延長, 最高速度, 車道幅員, 中央幅員, 交通量は, それぞれリンク延長, 規制最高速度, 車道幅員, 中央幅員, 平日の 24 時間交通量を意味する. 表-2 に示された結果から, 24 時間交通量は事故件数との間に高い相関をもち, いっぽう車線数と車道幅員は事故件数との間に弱い相関をもつことがわかる. ただし, 車線数, 車道幅員, 中央幅員の間には相互に強い相関がある。

次に, 説明変数と目的変数との間の関係を見るために, ツリーモデル¹⁹⁾ を使って分析を行った. 表-2 に示した説明変数の候補を使い, ツリーを計算した結果は図-3 に示したようになり, 24 時間交通量の値でリンクの事故件数が決定されていることがわかった. 他の変数はとり上げられなかった。

以下では, ポアソン回帰モデルの推定結果を次の順番で提示する. 推定はいずれも rstan を使って行った²⁰⁾. まず空間的自己相関を組み込まないポアソン回帰モデルの推定結果を示す. このモデルは通常の一般化線型回帰モデルと実質的に同等のモデルである. 次に空間的自己相関を組み込んだ Leroux *et al.* のモデルの推定結果を示し, 二つのモデルの比較を行う。

なお, McMC の計算では, チェーンを四つ使い, 一つのチェーンごとにシミュレーション回数を 12 万回,

ウォームアップ回数を 2 万回, *thinning* を 10 とした。また, 説明変数それぞれはスケーリングした上で計算に供している。

空間的自己相関を組み込まないポアソン回帰モデルの推定結果 (表-3) を見ると, 導入した説明変数のうちの二つ (車線数, 中央帯幅員) の 95 パーセント信用区間には 0 が含まれており, これらの変数は有意とはいえないことがわかる。

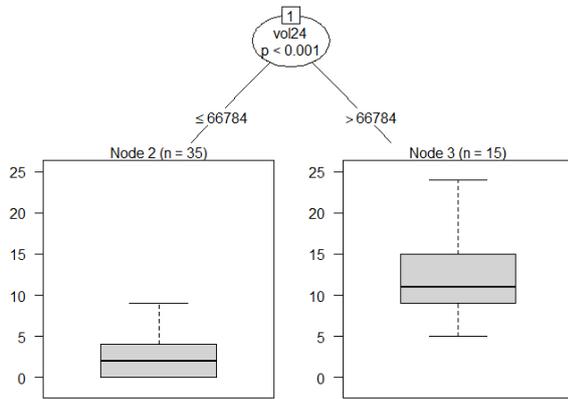


図-3 ツリーモデルの計算結果: 24 時間交通量の値が 66,784 台以下か, それより大きいかで二分割されている

表-3 空間的自己相関が組み込まれていないポアソン回帰モデルの推定結果

変数	2.5%	中央値	97.5%	\hat{R}
切片	2.43	3.34	4.21	1
車線数	-1.08	-0.42	0.23	1
最高速度	-2.71	-1.55	-0.39	1
中央帯幅員	-0.46	0.95	2.36	1

事後分布からのサンプルを使って各リンクの事故件数の値の中央値と 80 パーセント信用区間を求め, 観測された件数と比較した結果を図-4 に示した。なお, 観測された事故件数が同一のデータが複数存在するため, この図では件数の観測値に乱数を加えて, 点と線が互いに重ならないようにしてある。また, 路線別に描画に使う点の色と形状を変えている。

この図によると, 中央値は 45 度の直線 ($y = x$) から外れているケースが多く, とくに観測された事故件数の値が大きい場合には事後分布から求められた中央値は観測値に比べ過少になっていることがわかる。このことから, 空間的自己相関を組み込まないモデルは良い結果を与えてくれないと結論づけることができる。

次に空間的自己相関を組み込んだ Leroux *et al.* のモデルの推定結果を提示する。モデル推定の過程では様々な説明変数の組み合わせを試したが, 結果としては表-4 に示した, 説明変数として切片と 24 時間交通量のみを組み込んだモデルが選ばれた。この結果は図-3 に掲

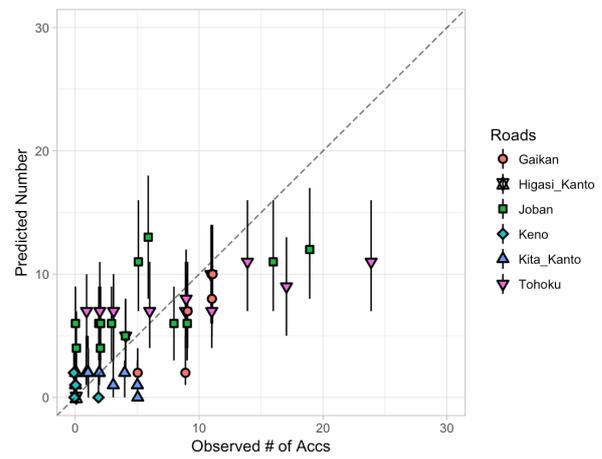


図-4 空間的自己相関が組み込まれていないポアソン回帰モデルの観測値と事後分布から求めた予測値の中央値と 80%信用区間 (縦の線分)

げたツリーモデルの結果とも符合している。表-4 から切片の 95 パーセント信用区間に 0 が含まれていることがわかる。交通量の係数は正值となっており, 交通量が増えるのに伴って事故件数も比例的に増加することが示されている。 ρ の中央値は 0.85 と 1 に近くリンクの事故件数の間には強い相関関係がある。

表-4 空間的自己相関が組み込まれたポアソン回帰モデルの推定結果

変数	2.5%	中央値	97.5%	\hat{R}
切片	-0.56	0.30	1.18	1.01
交通量	0.47	1.58	2.62	1.00
ν	0.80	1.55	2.98	1.00
ρ	0.41	0.85	0.99	1.00

(注) $\nu = 1/\tau^2$ であり, τ^2 は式 (2) に現れるものと同じ。

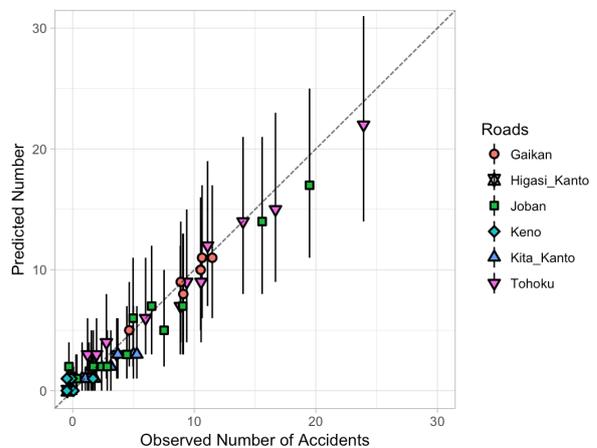


図-5 空間的自己相関が組み込まれたポアソン回帰モデルの観測値と事後分布から求めた予測値の中央値と 80%信用区間 (縦の線分)

各リンクの事故件数の予測値について中央値と 80 パーセント信用区間を求め、観測された件数と比較した結果を図-5 に示した。この図からは、予測値の中央値は 45 度の直線 ($y = x$) の近くに位置していることがわかり、空間的自己相関が組み込まれていないモデル (図-4) よりも当てはまりが改善されていることがわかる。ただし、80 パーセント信用区間は図-4 より広がっている。

5. 結論

本研究では、高速道路のリンク上で発生した交通事故件数を分析対象としてとり上げ、空間的な自己相関に着目して分析を行った。この分析によって得られた結論をとりまとめると次のとおりである。

まず、相互に連結されているリンクは隣接しているものとして隣接行列を定め、クラスタリングを表す指標であるモランの I、Geary の C および Getis-Ord の G を計算した。いずれの指標の値も、事故が多発しているリンクがクラスタリングしていることを示した。これまで事故分析に広く用いられてきたポアソン回帰モデル、負の二項分布モデル、ゼロ強調モデルでは、リンクの事故件数が独立であるという前提が置かれており、空間的な自己相関のあるデータをモデル化するには適切でないと考えられる。

この結果を踏まえて、空間的な自己相関を陽にとり込んだポアソン回帰モデルを McMC 推定した。このモデルではポアソン分布の期待値を表す線型予測子の中に条件付き自己回帰 (CAR) モデルにしたがうランダム効果項を含めた。CAR モデルとしては、これまで intrinsic CAR が広く用いられてきているが、このモデルは空間的自己相関が強い場合には適している一方で自己相関が弱い状態を表すことができないので、ここでは intrinsic CAR を一般化した Leroux *et al.* のモデルを使った。このモデルを推定したところ、高速道路の事故件数によく当てはまるモデルを得ることができた。

ただし、本研究では以下の点について十分な検討ができなかったことを述べておく必要がある。

まず、対象としたネットワークが小規模なものではない点である。これは、ネットワークが大きくなると隣接行列もそれに応じて大きな行列となり、McMC 計算における確率密度の計算に長時間を要するようになると見込まれたため、まず小規模なネットワークでの計算時間を把握することを目的としたためである。本文では詳しく触れなかったが、Joseph の計算法¹⁸⁾を用いると計算時間が大幅に短縮されることがわかったので、大規模なネットワークを対象にしたモデル推定を視野に入れて今後の分析を行うこととしたい。

また、高速道路のリンクでは事故の要因と考えられ

る変数が十分把握できないことも明らかになった。交通量をはじめとして、道路の線型 (曲率半径、勾配等) に関わるデータについても利用できるものは少なく、分析に支障をきたす虞がある。関係機関には事故に関わるデータの公開あるいは提供を検討していただきたい。

謝辞: 本報告は、2016 年に公益財団法人交通事故総合分析センターと国立大学法人東京大学の間で実施された共同研究の成果の一部である。研究に携わった方々、特に事故データの集計を担当された北野さんに感謝する。

参考文献

- 1) 山田晴利, Teerayut, H., 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, 第 49 巻, 9 ページ, 2014.
- 2) 久保拓弥: データ解析のための統計モデリング入門, 岩波書店, 2012.
- 3) Faraway, J. J.: *Extending the linear model with R: generalized linear, mixed effects and non-parametric regression models*, Chapman and Hall/CRC, 2005.
- 4) Sharma, A. K. and Landge, V. S.: Zero inflated negative binomial for modeling heavy vehicle crash rate on Indian rural highway, *International Journal of Advances in Engineering & Technology*, vol. 5, issue 2, pp. 292–301, 2013.
- 5) Duddu, V. R. and Pulugurtha, S. S.: Neural networks to estimate crashes at zonal level for transportation planning, presented at European Transport Conference 2012.
- 6) Paweenuch, S., Yamada, H., Teerayut, H. and Shibasaki, R.: Traffic accidents risk analysis based on road and land use factors using GLMs and zero-inflated models, *Proceedings of CUPUM 2015*, paper no. 320, 26 pages, 2015.
- 7) Brunson, C. and Comber, L.: *An introduction to R for spatial analysis and mapping*, Sage Publishers Inc., 2015.
- 8) Banerjee, S., Bradley, P. and Gelfand, A. E.: *Hierarchical modeling and analysis for spatial data*, 1st edition, Taylor & Francis, 2003.
- 9) 岩波データサイエンス刊行委員会 (編): ベイズ推論と MCMC のフリーソフト, 岩波書店, 2015 年.
- 10) Lee, D.: A comparison of conditional autoregressive models used in Bayesian disease mapping, *Spatial and Spatio-Temporal Epidemiology*, vol. 2, pp. 79–89, 2011.
- 11) Leroux, B., Lei, X. and Breslow, N: Estimation of disease rates in small areas: a new mixed model for spatial dependence in Halloran, M. and Berry, D. edited *Statistical models in epidemiology, the environment and clinical trials*, pp. 135–178, Springer-Verlag, 1999.
- 12) 金杉浩, 黒川茂莉, 松村茂樹, 関本義秀: 携帯電話基地局通信情報の行動分析への適用可能性, 第 32 回交通工学研究発表会論文集, pp. 309–315, 2012 年.
- 13) 長谷川瑠子, 関本義秀, 金杉浩, 新井亜弓, Teerayut, H., Apichon, W.: CDR データを用いたダッカの人の流動把握, 第 22 回地理情報システム学会講演論文集, 第 22 巻, 2013 年.
- 14) 関本義秀, 檜山武浩, 長谷川瑠子, 金杉浩: スパースな携帯電話通話履歴を用いたリンク交通量の推定—ダッカの事例, *交通工学論文集*, vol. 1, no. 4, pp. A1–A8, 2013 年.
- 15) 平成 22 年度道路交通センサス一般交通量調査 DVD-ROM, 一般社団法人交通工学研究会, 2012 年.
- 16) R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna,

Austria.

- 17) Stan Development Team: *Stan modeling language users guide and reference manual*, Stan version 2.14.0, 2016.
- 18) Joseph M: Exact sparse CAR models in STAN, <http://mc-stan.org/documentation/case-studies/mbjoseph-CARStan.html> <2017 年 4 月 2 日アクセス>
- 19) Crawley, M. J.: *Statistics: an introduction using R*, 1st edition, Wiley, 2005.
- 20) 松浦健太郎: Stan と R でベイズ統計モデリング, 共立出版, 2016 年.

(2017. 4. 28 受付)

A STUDY ON THE NUMBER OF TRAFFIC ACCIDENTS ON EXPRESSWAY LINKS

Harutoshi YAMADA, Quanjin Chen, Yoshio TANAKA and Ryosuke SHIBASAKI

In analyzing traffic accidents data, it is well known that the number of road links where no traffic accidents occurred is in the majority. This makes the building of accidents models difficult. To solve this problem, negative binomial regression models or zero inflated models have been conventionally used, but their results have been unsatisfactory because geographical autoregression is not considered in them. In this paper, new models of the number of traffic accidents on expressways have been developed introducing a random effect variable that follows a Conditional AutoRegressive (CAR) distribution into the link function that expresses the expectation of a Poisson distribution. It has been found that this model gives us an acceptable result. In particular, the CAR model developed by Leroux *et al.* has been proved to be more satisfactory than the intrinsic CAR model that has been used widely thus far.