研究論文

交通量変動を考慮した新型コロナウイルス感染症の流行期における都道府県別自動車事故頻度の分析

谷川 祥太* 野村 俊一† 佐野 誠一郎‡

2023 年 10 月 7 日投稿
2024 年 2 月 8 日受理

概要

新型コロナウイルス感染症は，人々に対して深刻な被害を及ぼしただけでなく，社会生活の大きな変革をもたらしてきた．特に，緊急事態宣言による移動制限やテレワークの普及は，交通量の大幅な減少を引き起こし，それに伴い交通事故の件数も顕著に減少した．その後，ワクチン接種やウイルス変異に伴って行動制限は徐々に緩和され，交通量も次第に元の水準に向けて回復してきた．このため損害保険会社にとっては，料率の算定や収支管理の観点から，交通量の回復に伴うクレーム頻度の変動を正確に予測することが重要となる．

そこで本研究では，自動車損害賠償責任保険の都道府県別支払件数の 2013～21 年度推移を時空間ベイズモデルにより分析する．都道府県間のクレーム頻度較差を推定するとともに，自動車走行距離を共変量に導入することで交通量の減少がクレーム頻度に与えた影響を議論し，さらに交通量が回復した 2022 年度のクレーム頻度予測を行う．

キーワード：新型コロナウイルス感染症，自動車損害賠償責任保険，クレーム頻度分析，時空間ベイズモデル

* 早稲田大学商学学術院大学院会計研究科　〒169-8050 東京都新宿区西早稲田 1-6-1　Email: shota.tanikawa@toki.waseda.jp
† 早稲田大学商学学術院大学院会計研究科　〒169-8050 東京都新宿区西早稲田 1-6-1　Email: snomura5@aoni.waseda.jp
‡ 共栄火災海上保険株式会社　〒105-8604 東京都港区新橋 1-18-6　Email: se.sano@kyoeikasai.co.jp　本稿の内容はすべて筆者の個人的な見解であり，筆者が所属する団体の見解とは無関係である．

1 はじめに

新型コロナウイルス感染症（COVID-19）の蔓延は，近年最も大きな公衆衛生上の危機として世界を席巻してきた．この感染症は，数多くの命を奪うだけでなく，我々の日常生活やビジネス活動にも大きな変化をもたらしてきた．緊急事態宣言や感染拡大防止策としてのテレワークの導入により公共の場での人の動きや交通量は大幅に減少した結果，交通事故の発生率も顕著に低下した（島田[2021]）．

しかし，ワクチン接種の普及や新型コロナウイルスの変異株への置き換わりなどによって，新型コロナウイルス感染症による死亡率は低下し，それに伴って社会の活動も徐々に正常化してきた．行動制限により大幅に減少していた交通量も次第に回復を見せており，大きく減少していた交通事故件数も今後は増加に転ずる可能性がある．こうした変化は，自動車保険が大きなウェイトを占めている損害保険会社の業績や収支バランスにも影響を及ぼしうる．交通量の変動はクレームの頻度や性質に変化をもたらすため，これを的確に予測して保険料率へと反映させることが，損害保険業界にとっての大きな課題となっている．

そこで本稿では，一般に公開されている自動車損害賠償責任保険の都道府県別のクレーム頻度を題材として分析を行う．尚，自動車損害賠償責任保険は対人賠償責任のみ補償する強制加入の保険であるが，自動車保険は自動車事故による損害全般を扱う任意保険であり，対人賠償補償だけではなく対物賠償補償や車両補償など補償範囲を自由に選ぶことができ，更に免責金額の設定に応じて請求されない事故が存在するため，そうした違いを無視して精緻な分析を行うことは困難である．自動車損害賠償責任保険と任意の自動車保険ではこのように被保険者集団や補償内容が異なることなどにより，両者の事故の発生動向は必ずしも一致しないが，任意保険について分析に利用できる十分な公開データがないことから本研究では自動車損害賠償責任保険のデータを用いた．交通量の変化がクレーム頻度にもたらす影響を定量的に分析するとともに，交通量の回復によってクレーム頻度がどのように変化するか予測を行う．クレーム頻度の推移には，交通量の変化だけでなく，自動車の車両安全装置の普及などに伴う長期的な減少トレンドや，都道府県間の較差も見られるため，本研究ではベイズ時空間モデルを用いた分析を試みる．

本研究で扱うベイズ時空間モデルは，一般化線形モデルに時間，空間あるいは双方に依存するランダム効果を加えることによりデータの持つ時空間的な相関構造を取り入れた階層ベイズモデルである．時間依存のランダム効果にはトレンドを考慮した時系列モデルが仮定され，空間依存のランダム効果には隣接地域の値に間の従属性を持つ条件付き自己回帰(Conditional Autoregressive; CAR)モデルが仮定される．空間ランダム効果を取り入れたベイズ空間モデルは損害保険の地域別クレーム分析に利用されており（Assuncao et al.[2014]），Boskov and Verrall[1994]は CAR モデルを用いたベイズ空間モデルにより郵便番号で区分けされた地域別の損害率を推定している．Gschlößl and Czado[2007]は地域別のクレーム頻度と規模のそれぞれをベイズ空間モデルにより推定し，Tufvesson et al.[2019]はさらにスパース正則化による共変量選択を取り入れた．ベイズ時空間モデルについては，Ozaki et al.[2009]がブラジルにおけるトウモロコシの地域別収量予測に基づく農作物保険の料率算定へと応用しており，佐野[2022]は国内の都道府県別がん死亡数予測への応用例を示している．こうした階層ベイズモデルでは従来からマルコフ連鎖モンテカルロ法に基づく推論が行われてきたが，Assuncao et al.[2014]，Tufvesson et al.[2019]および佐野[2022]では Integrated Nested Laplace Approximation(INLA)を用いた推論が行われている．Rue et al.[2009]により提案された INLA は，ランダム効果などの潜在変数に対するベイズ推論をラプラス近似を用いて高速にかつ高精度に実現できる近似推定手法であり，R の INLA パッケージにはベイズ時空間モデルを手軽に実装する様々な関数が用意されている．

本研究では，佐野[2022]の分析で用いられた 2 次のランダムウォークに従う時間ランダム効果，Leroux モデル

（Leroux et al.[2000]）に従う空間ランダム効果，独立同一分布に従う時空間ランダム効果を取り入れた時空間ベイズモデルを検討し，R の INLA パッケージを利用して推論を行う．Tufvesson et al.[2019]は車両保険のクレーム頻度と規模の空間モデル分析に人口・経済に関する膨大な共変量を検討しているが，本研究の特徴は自動車走行距離の都道府県別集計値を共変量に導入した点にある．また，新型コロナウイルス感染拡大に伴う移動制限によって交通事故の発生頻度は大きく減少しており，釜賀[2022]はコロナ禍における各都道府県の交通事故死者数の変化のメカニズムについて考察し，緊急事態宣言期間(日)を変数に考慮した決定木（ID3）にて交通事故死者数の推移(増加・減少)を予測している．兵頭[2022]は都市高速道路における事故データを用いて新型コロナウイルスの感染拡大前後を時期ダミー変数としたポアソン回帰モデルによって事故類型別の分析を行った．これらの先行研究に対し，本研究ではクレーム頻度のここ数年続く減少トレンドや都道府県間較差を時空間モデルとして織り込んだ上で自動車走行距離がクレーム頻度に与える影響を定量的に分析し，交通量回復を踏まえたクレーム頻度の将来予測を可能としたことが主な貢献となる．

本稿の構成は以下の通りである．次の 2 節では，使用するデータについて説明する．続く 3 節では時空間モデルについて詳述し，4 節でモデルの候補と INLA を用いた分析手法を示す．そして，5 節にてモデルによる推定結果と交通量変化を踏まえた予測結果を示し，最後に 6 節でまとめを述べる．

2 使用するデータ

本稿の分析には，損害保険料率算出機構の公表する損害保険料率算出機構統計集のうち「自賠責保険統計」及び，政府統計の総合窓口にて公表されている交通関係統計資料「自動車燃料消費量調査」を用いる．

「自賠責保険統計」では自動車損害賠償責任保険のクレームデータについて車種別，用途別，そして都道府県別に契約および支払の集計がされており，各地域の特徴を捉えることができる．本稿では，自動車損害賠償責任保険の基準料率において「離島以外の地域（沖縄県を除く）」の区分に属する 46 都道府県における車種「自家用乗用自動車」の新契約台数，支払件数（内.死亡を除いた件数）を用いて分析を行っていく．なお，支払件数のうち死亡件数を除いたものを 2.1 節にて説明する経過台数で除したものをクレーム頻度と定義する．

また，「自動車燃料消費量調査」では自動車の走行距離について燃料別，都道府県別，車種別の集計がされている．本稿では，「自動車燃料消費量調査」のうち自動車損害賠償責任保険の車種「自家用乗用自動車」に該当する図 1 赤枠の燃料・車種の区分における「走行距離」（km）の都道府県別の合計値を分析に使用する．

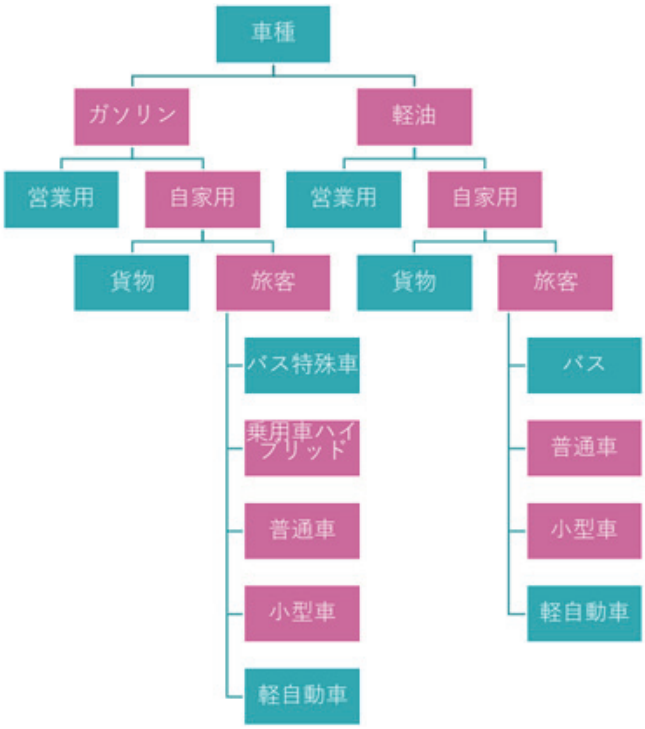

図１　自動車燃料消費量調査より使用する車種

本稿の分析期間は2013年度から2022年度までとする．ただし，「自賠責保険統計」は本稿執筆時点（2023年9月末）において損害保険料率算出機構が公開している2013年度から2021年度までのデータを用いる．図２には分析対象地域（沖縄県を除いた46都道府県）全体でのクレーム頻度と走行距離の指数（2013年度を１とおいた）を示している．新型コロナウイルスによる行動自粛の影響で2020年以降，走行距離の低下がみられる．それに伴いクレーム頻度も減少しているが，図２灰色線で示した両者の比（クレーム頻度の指数÷走行距離の指数）を見ると，走行距離の減少を加味した上でもなおクレーム頻度は2015年頃から長期的な減少トレンドにあることが見てとれる．これは，自動車の安全運転装置などの車両安全対策の普及が背景にあるものとみられる．

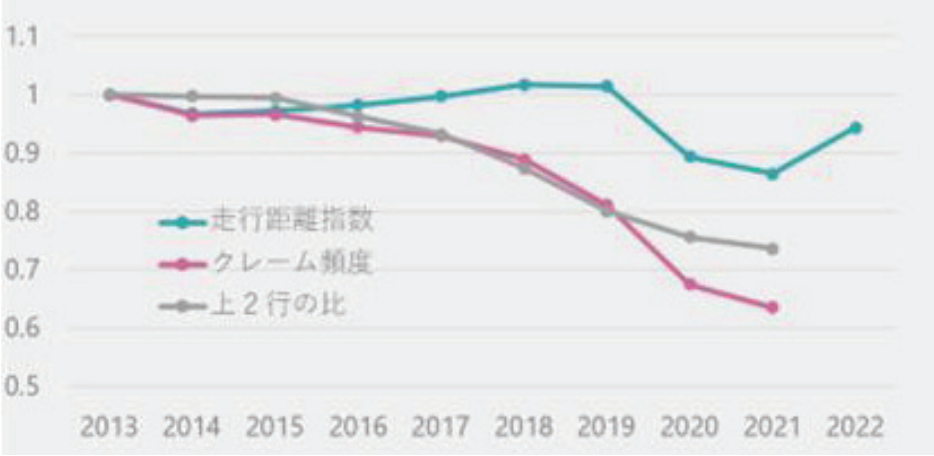

図２　走行距離およびクレーム頻度の指数（2013年度比）の推移

2.1 経過台数

「自賠責保険統計」では年間の新契約台数を把握しており，クレーム頻度を算出するにあたっては各年度における経過台数を求めておく必要がある．本稿では新契約が年度内で一様に発生しているとする仮定に基づいて，次の方法により経過台数を求める．i年度における保険期間jか月の新契約台数を$a_{i,j}$とおく．これら新契約の保険始期がi年度内で一様に分布しているものと仮定することで，当該新契約 1 台あたりのt年度における経過台数$p_{i,j,t}$（単位：台×年）が次式によって求まる．

$$p_{i,j,t} = \begin{cases} \frac{1}{2} - \frac{1}{2}\left(\max\left\{1 - \frac{j}{12}, 0\right\}\right)^2 & t = i \\ 1 - \frac{1}{2}\left(\max\left\{1 - \frac{j}{12} - i + t, 0\right\}\right)^2 & i < t \leq i + \frac{j}{12} \\ \frac{1}{2}\left(i + \frac{j}{12} + 1 - t\right)^2 & i + \frac{j}{12} < t < i + \frac{j}{12} + 1 \\ 0 & \text{上記以外} \end{cases}$$

さらに，i年度における都道府県kの新契約台数を$b_{i,k}$とおく．このとき，t年度における都道府県kの経過台数$E_{t,k}$は次式によって求まる．

$$E_{t,k} = \sum_{i=t-4}^{t} \sum_{j=1}^{37} \frac{a_{i,j}}{\sum_{j=1}^{37} a_{i,j}} * b_{i,k} * p_{i,j,t}.$$

2021 年度における経過台数を計算するには 2021 年度における保険期間 1 - 37 か月，2020 年度保険期間 1 - 37 か月，2019 年度保険期間 13 - 37 か月，2018 年度保険期間 25 - 37 か月，2017 年度保険期間 37 か月の新契約台数が必要になる．具体例として 2018 年度の新契約のうち保険期間６か月と 37 か月の新契約 1 台あたりの各年度における経過台数を求める．図３および図４は経過台数のイメージであり，各年度内に含まれる面積が求める経過台数になる．

2018 年度における保険期間６か月の新契約 1 台あたりの 2018，2019，2020，2021，2022 年度における経過台数は次の通り．

$$p_{2018,6,2018} = \frac{1}{2} - \frac{1}{2}\left(\max\left\{1 - \frac{6}{12}, 0\right\}\right)^2 = \frac{3}{8},$$

$$p_{2018,6,2019} = \frac{1}{2}\left(2018 + \frac{6}{12} + 1 - 2019\right)^2 = \frac{1}{8},$$

$$p_{2018,6,2020} = P_{2018,6,2021} = P_{2018,6,2022} = 0.$$

2018 年度における保険期間 37 か月の新契約 1 台あたりの 2018，2019，2020，2021，2022 年度における経過台数は次の通り．

$$p_{2018,37,2018} = \frac{1}{2} - \frac{1}{2}\left(\max\left\{1 - \frac{12}{12}, 0\right\}\right)^2 = \frac{1}{2},$$

$$p_{2018,37,2019} = 1 - \frac{1}{2}\left(\max\left\{1 - \frac{37}{12} - 2018 + 2019, 0\right\}\right)^2 = 1,$$

$$p_{2018,37,2020} = 1 - \frac{1}{2}\left(\max\left\{1 - \frac{37}{12} - 2018 + 2020, 0\right\}\right)^2 = 1,$$

$$p_{2018,37,2021} = 1 - \frac{1}{2}\left(\max\left\{1 - \frac{37}{12} - 2018 + 2021, 0\right\}\right)^2 = \frac{167}{288},$$

$$p_{2018,37,2022} = \frac{1}{2}\left(2018 + \frac{37}{12} + 1 - 2022\right)^2 = \frac{1}{288}.$$

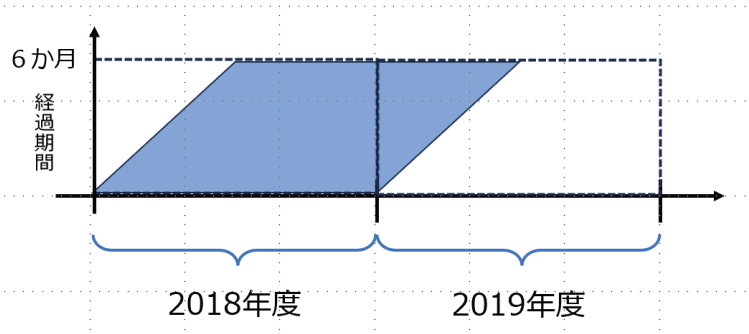

図３　2018 年度における保険期間６か月の新契約１台あたりの経過台数のイメージ

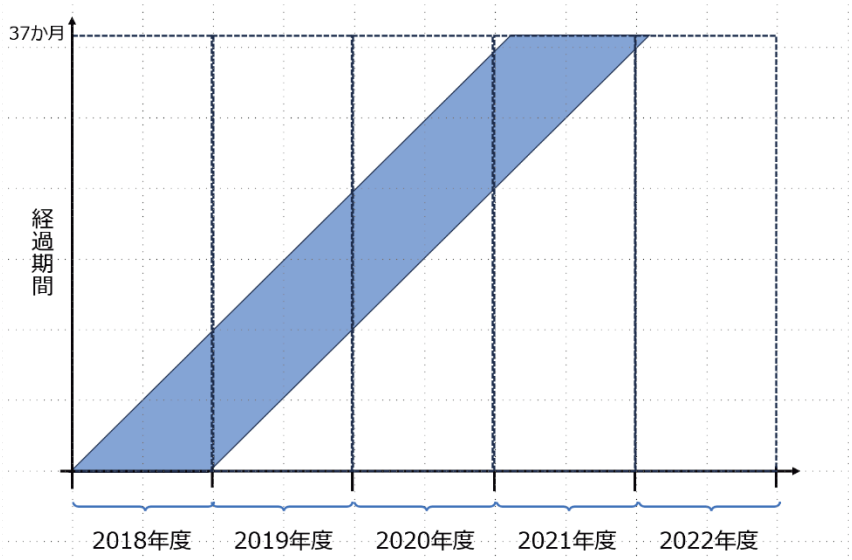

図４　2018 年度における保険期間 37 か月の新契約１台あたりの経過台数のイメージ

3　時空間ベイズモデル

本節では，本研究の分析に用いる時空間ベイズモデルを導入する．そのベースとなるモデルとして，t 年度目の k 番目の地域における支払件数$y_{t,k}(t=1,\dots,T,k=1,\dots K)$がそれぞれ独立に平均$\lambda_{t,k}$のポアソン分布$\mathrm{Po}(\lambda_{t,k})$に従うことを仮定した次式の一般化線形モデル（generalized linear model; GLM）を考える：

$$y_{t,k} \sim \mathrm{Po}(\lambda_{t,k}), \tag{1}$$

$$g(\lambda_{t,k}) = \log \lambda_{t,k} = \eta_{t,k}, \tag{2}$$

$$\eta_{t,k} = \beta_0 + \sum_{i=1}^{p} \beta_i x_{i,t,k} + \log E_{t,k}\,. \tag{3}$$

ここでは，上式の通りリンク関数gを対数リンクとしており，支払件数$y_{t,k}$に対して p 種の共変量$x_{1,t,k},\dots,x_{p,t,k}$を持つものとする．また，$E_{t,k}$は支払件数$y_{t,k}$に対応する経過台数であり，クレーム頻度$\lambda_{t,k}$が経過台数$E_{t,k}$に比例するようにオフセット項 $\log E_{t,k}$（回帰係数が掛からない固定加算項）を導入している．

本稿では上式の GLM に，以降で説明する時間依存，空間依存，時空間依存のランダム効果を導入することにより，時空間ベイズモデルへと拡張する．これら全てのランダム効果を導入したとき，GLM の式(3)は次式のように変わる：

$$\eta_{t,k} = \beta_0 + \sum_{i=1}^{p} \beta_i x_{i,t,k} + \log E_{t,k} + \gamma_t + u_k + \psi_{t,k}. \tag{4}$$

ここで，各ランダム効果にはそれぞれ以降で紹介する事前分布を与え，さらに事前分布のもつ超パラメータにも超事前分布を導入することで階層ベイズモデルへと拡張する．さらに，回帰係数$\boldsymbol{\beta}=\{\beta_0,\beta_1,\dots,\beta_p\}$にもそれぞれ事前分布$\mathrm{N}(0,1000)$を与えることで全てのパラメータに事前分布が課されたフルベイズモデルとする．

3.1 時間依存のランダム効果

まず，クレーム頻度の時間的変化を捉えるために時間依存のランダム効果$\boldsymbol{\gamma}=\{\gamma_1,\gamma_2,\cdots,\gamma_T\}$を導入する．図２から観察されるように，クレーム頻度を走行距離で割った指数は2015年以降減少トレンドにあり，その傾きは緩やかに変化している．そこで，トレンドを前年からある程度保ちながら時間推移するモデルとして，$t=3,\dots,T$において次の２次のランダムウォークモデルを考える：

$$\Delta^2\gamma_t=\Delta\gamma_t-\Delta\gamma_{t-1}=(\gamma_t-\gamma_{t-1})-(\gamma_{t-1}-\gamma_{t-2})\sim\mathrm{N}\left(0,\tau_\gamma^{-1}\right) \tag{5}$$

ただし，Δ,Δ^2は１階，２階の階差オペレータであり，$\Delta\gamma_t=\gamma_t-\gamma_{t-1}$はランダム効果の１時点間の変化量すなわち傾きを表す．ゆえに，上式はランダム効果の傾き$\Delta\gamma_t$が，正規分布$\mathrm{N}\left(0,\tau_\gamma^{-1}\right)$に従う変化量が毎時点加わることで少しずつ変化するモデルとなっている．このモデルの利点は，直近の傾きすなわちトレンドに従って将来予測が行われる点にある．このことは，将来時点$t=T+1,T+2,\dots$におけるランダム効果の傾き$\Delta\gamma_t$が，直近時点の傾き$\Delta\gamma_T$と平均０の２階階差$\Delta^2\gamma_{T+1},\cdots,\Delta^2\gamma_t$との和により

$$\Delta\gamma_t=\Delta\gamma_T+(\Delta\gamma_{T+1}-\Delta\gamma_T)+\cdots+(\Delta\gamma_t-\Delta\gamma_{t-1})=\Delta\gamma_T+\Delta^2\gamma_{T+1}+\cdots+\Delta^2\gamma_t$$

と表されるため，$\gamma_1,\dots,\gamma_T$が所与の下でその条件付き平均は$E(\Delta\gamma_t|\gamma_1,\dots,\gamma_T)=\Delta\gamma_T$となり，ゆえに将来時点のランダム効果の平均すなわち予測値は

$$\begin{aligned}E(\gamma_t|\gamma_1,\dots,\gamma_T)&=E(\gamma_T+\Delta\gamma_{T+1}+\cdots+\Delta\gamma_t|\gamma_1,\dots,\gamma_T)=\gamma_T+E(\Delta\gamma_{T+1}|\gamma_1,\dots,\gamma_T)+\cdots+E(\Delta\gamma_t|\gamma_1,\dots,\gamma_T)\\&=\gamma_T+(t-T)\Delta\gamma_T\end{aligned} \tag{6}$$

と直近のγ_Tの値から傾き$\Delta\gamma_T$で増減し続けることから分かる．なお，式(5)に現れる事前分布$\mathrm{N}\left(0,\tau_\gamma^{-1}\right)$の精度（分散の逆数）を表す超パラメータ$\tau_\gamma$は，超事前分布としてガンマ分布$\Gamma(1,0.0005)$に従うこととする．

また，もう一つの選択肢として，時間依存のランダム効果を用いる代わりに，時点tを共変量として一定の傾きで作用する項$\beta_{year}t$を式(4)第２項の中に組み入れたモデルも，後の解析のモデル候補として考える．

3.2 空間依存のランダム効果

本研究で扱うような地域毎に集計されたデータでは，隣接する地域どうしの観測値が一定の相関関係をもつことが多い．たとえば，ある地域で支払件数が多ければ，その隣接地域でも支払件数は多い傾向にあることが予想される．そのような空間相関構造をモデルに取り込むための方策として，条件付き自己回帰（conditional auto-regressive; CAR）モデル(Besag[1974])に従う空間依存のランダム効果$\boldsymbol{u}=\{u_1,u_2,\dots,u_K\}$を導入する．

各地域$k=1,\cdots,K$について，その隣接地域の集合$N(k)\subseteq\{1,2,\dots,K\}$が定義され，地域$k$を除く全地域における$\{u_j;j\neq k\}$の値が与えられた条件下で，$u_k$の従う条件付き分布が隣接地域の値$\{u_j;j\in N(k)\}$にのみ依存してそれ以外の地域には依存しないこと（条件付き独立性）を仮定したモデルを一般にマルコフ確率場という．特に，u_kが隣接地域の値$\{u_j,j\in N(k)\}$の線形和で表される条件付き平均と，地域kに依存する条件付き分散をもつ正規分布に従うモデルは，CARモデルと呼ばれる．

CARモデルにはいくつかのバリエーションが存在する．最も代表的なものは，u_kに対する条件付き分布として次の正規分布を仮定したICAR（intrinsic CAR)モデルである：

$$u_k|\{u_j; j \neq k\} \sim \mathrm{N}\left(\frac{\sum_j w_{kj} u_j}{\sum_j w_{kj}}, \frac{\tau_u^{-1}}{\sum_j w_{kj}}\right)$$

ただし，w_{kj}は地域kに対し地域jが隣接する場合は1，そうでない場合は0をとる．そのため，u_kの条件付き平均は隣接地域の値$\{u_j; j \in N(k)\}$の平均値となっており，さらに，隣接地域数が多いほど条件付き分散は小さくなるため隣接地域の値による束縛が大きくなることが分かる．

一方で，ICARモデルはu_kの条件付き平均が隣接地域の値に強く依存しており，より弱い空間相関を表現できるモデルがその後いくつか提案されてきた．Besag et al.[1991]はICARモデルの条件付き平均に，それぞれ独立同一の正規分布に従うノイズを加えることで空間相関を弱めたBesag-York-Mollie(BYM)モデルを提案し，Stern and Cressie[1999]はICARモデルの条件付き平均に対して区間(0,1)上の一様分布に従う超パラメータを掛けることで空間相関を弱めたモデルを提案している．本稿では，ICARモデルの条件付き分布を次式のように改変したLerouxモデル(Leroux et al.[2000])を利用する：

$$u_k|\{u_j, j \neq k\} \sim \mathrm{N}\left(\frac{\rho \sum_j w_{kj} u_j}{\rho \sum_j w_{kj} + 1 - \rho}, \frac{\tau_u^{-1}}{\rho \sum_j w_{kj} + 1 - \rho}\right) \tag{7}$$

ここで，ρは$0 \leq \rho \leq 1$の範囲の値をとる空間相関の強さを調整するパラメータであり，$\rho = 1$の場合にはICARモデルとなり，$\rho = 0$の場合には$u_1, u_2, \ldots, u_K$は互いに独立で無相関となる．Lerouxモデルは理論的にも実用上においても最良であるという指摘(Lee[2011])もあり，今日において特に広く用いられている．式(7)に現れる2つの超パラメータτ_u（全体の分散を調整する超パラメータ）とρ（空間相関の強さを調整するパラメータ）に対してはそれぞれ超事前分布として，$\tau_u \sim \Gamma(1, 0.0005), \log\frac{\rho}{1-\rho} \sim \mathrm{N}(0, 0.45^{-1})$を与える．

なお，Lerouxモデルは式(7)の条件付き分布をもって定義されているが，Brook’s lemmaにより$u_1, u_2, \ldots, u_K$全体としては多変量正規分布に従うことが保証されている（たとえば村上[2022]第10章）．

3.3 時空間依存のランダム効果（時間と空間の交互作用項）

最後の時空間依存のランダム効果$\boldsymbol{\psi} = \{\psi_{t,k}; t = 1, \ldots, T, k = 1, \ldots K\}$は，それぞれが独立に同じ正規分布$\mathrm{N}(0, \tau_\psi^{-1})$に従うものとして定義する．これは，前述した時間依存のランダム効果と空間依存のランダム効果のみでは捉えきれない時間と空間の交互作用項と見なされる．また，その精度を表す超パラメータτ_ψについては，超事前分布としてガンマ分布$\Gamma(2, 0.5)$に従うものとする．この超事前分布は，時間依存ならびに空間依存のランダム効果の精度パラメータに対する超事前分布$\tau_\gamma, \tau_u \sim \Gamma(1, 0.0005)$に比べて大きい値（すなわち小さい分散）をとりやすいよう設定した．このことにより，データの持つ純粋な時間的変動ならびに空間的変動が，分散の小さい時空間依存のランダム効果よりも，分散の大きい時間依存・空間依存のランダム効果に寄せられて推定されることになる．

支払件数のようなカウントデータに対しては式(1)のようにポアソン分布を用いたモデルが当てはめられることが多い．ポアソン分布は少数の法則により導かれるカウントデータの分布であり，平均と分散が等しいという特徴をもつ．しかし現実のカウントデータでは，分散が平均を上回ることが多く，これを過分散という．ここでのランダム効果$\psi_{t,k}$は，観測値$y_{t,k}(t = 1, \ldots, T, k = 1, \ldots K)$ごとに独立に異なる値をとり，ポアソン分布$\mathrm{Po}(\lambda_{t,k})$の平均$\lambda_{t,k}$をランダムに撹乱させるため，過分散にも適合できるよう観測値$y_{t,k}$のばらつきを増大させる役割をもっている．実際，時空間ランダム効果$\boldsymbol{\psi}$を除く全てのパラメータ$\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{u}$の値が特定された条件の下で，各観測値$y_{t,k}(t = 1, \ldots, T, k = 1, \ldots K)$は互いに独立であり，さらに

$$\mu_{t,k} = \eta_{t,k} - \psi_{t,k} = \beta_0 + \sum_{i=1}^{p} \beta_i x_{i,t,k} + \log E_{t,k} + \gamma_t + u_k$$

とおくと観測値$y_{t,k}$の平均は

$$
\begin{aligned}
E(y_{t,k}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{u}) &= E\left[E(y_{t,k}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{u},\boldsymbol{\psi})|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{u}\right] = E(\lambda_{t,k}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{u}) \\
&= E(e^{\eta_{t,k}}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{u}) = E(e^{\mu_{t,k}+\psi_{t,k}}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{u}) \\
&= e^{\mu_{t,k}}E(e^{\psi_{t,k}})
\end{aligned}
$$

となる一方で，分散は

$$
\begin{aligned}
V(y_{t,k}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{u}) &= E\left[V(y_{t,k}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{u},\boldsymbol{\psi})|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{u}\right] + V\left[E(y_{t,k}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{u},\boldsymbol{\psi})|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{u}\right] \\
&= E(\lambda_{t,k}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{u}) + V(\lambda_{t,k}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{u}) \\
&= e^{\mu_{t,k}}E(e^{\psi_{t,k}}) + e^{2\mu_{t,k}}V(e^{\psi_{t,k}}) > E(y_{t,k}|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{u})
\end{aligned}
$$

となるので，実質的に観測値の分散は平均を上回ることになる．

このような時空間依存のランダム効果は，残差分析においても重要な役割を果たす．もし，$\psi_{t,k}$の推定値の時空間的挙動に何らかの系統的な特徴が見られるならば，それは$\psi_{t,k}$が互いに独立であるとする仮定が崩れていることを意味し，データの時空間的挙動に関する新たな知見を得ることができる．

4 分析方法

4.1 モデル候補

本稿では2節に示したデータについて，解析期間を2013年度から2021年度（$t = 1, \ldots, 9$），解析対象地域は沖縄県を除く46都道府県（$k = 1, \ldots 46$）として，支払件数$y_{t,k}$を目的変数とした3節の時空間ベイズモデル（式(1)，(2)および(4)）により分析する．ただし，式(4)に現れる経過台数$E_{t,k}$は2.1節に示した方法により算出されたものとし，また共変量の候補として，時点tを共変量として一定の傾きで作用する時間トレンド項$\beta_{year}t$と，走行距離指数$x_{t,k}$（2013年度の各都道府県の走行距離を1とした時の各年度における都道府県ごとの比率）の対数値$\log x_{t,k}$を共変量とした項$\beta_{mileage}\log x_{t,k}$を検討する．さらに，3.1節から3.3節にて導入した時間依存，空間依存，時空間依存（時空間交互作用）のランダム効果についてもそれぞれモデルへの要否を検討するために，以下の各モデルについて解析を行い，後述するモデル選択基準を用いて適切なモデル選択を行う．

・GLM：時間トレンド項と都道府県ごとの回帰係数β_kを設けた通常の一般化線形モデル

$$\eta_{t,k} = \log E_{t,k} + \beta_0 + \beta_{year}t + \beta_k$$

・Model 1:空間ランダム効果+時間ランダム効果

$$\eta_{t,k} = \log E_{t,k} + \beta_0 + u_k + \gamma_t$$

・Model 2:空間ランダム効果+時間ランダム効果+時空間交互作用

$$\eta_{t,k} = \log E_{t,k} + \beta_0 + u_k + \gamma_t + \psi_{t,k}$$

・Model 3:空間ランダム効果+時間トレンド項

$$\eta_{t,k} = \log E_{t,k} + \beta_0 + u_k + \beta_{year}t$$

・Model 4:空間ランダム効果+時間トレンド項+時空間交互作用

$$\eta_{t,k} = \log E_{t,k} + \beta_0 + u_k + \beta_{year}t + \psi_{t,k}$$

・Model 5:Model 1+走行距離指数

$$\eta_{t,k} = \log E_{t,k} + \beta_0 + u_k + \gamma_t + \beta_{mileage}\log x_{t,k}$$

・Model 6:Model 2+走行距離指数

$$\eta_{t,k} = \log E_{t,k} + \beta_0 + u_k + \gamma_t + \psi_{t,k} + \beta_{mileage}\log x_{t,k}$$

・Model 7:Model 3+走行距離指数

$$\eta_{t,k} = \log E_{t,k} + \beta_0 + u_k + \beta_{year} t + \beta_{mileage} \log x_{t,k}$$

・Model 8:Model 4+走行距離指数

$$\eta_{t,k} = \log E_{t,k} + \beta_0 + u_k + \beta_{year} t + \psi_{t,k} + \beta_{mileage} \log x_{t,k}$$

なお，空間ランダム効果u_kの従うLeroux モデルにおける都道府県間の隣接関係$w_{k,j}$（地域kに対し地域jが隣接する場合は$w_{k,j} = 1$，そうでない場合は$w_{k,j} = 0$をとる）については，地理情報システム（Geographic Information System：GIS）のデータを用いて設定した．具体的には，「国土数値情報（行政区域データ）」（国土交通省）から取得した都道府県別のシェイプファイルにおいて陸続きに隣接している都道府県間に加え，陸続きではないが北海道と青森県，兵庫県と徳島県，愛媛県と広島県，岡山県と香川県，山口県と福岡県の組も隣接しているものとして定義した．図５に隣接する都道府県間を線で結んだ隣接関係図を示す．

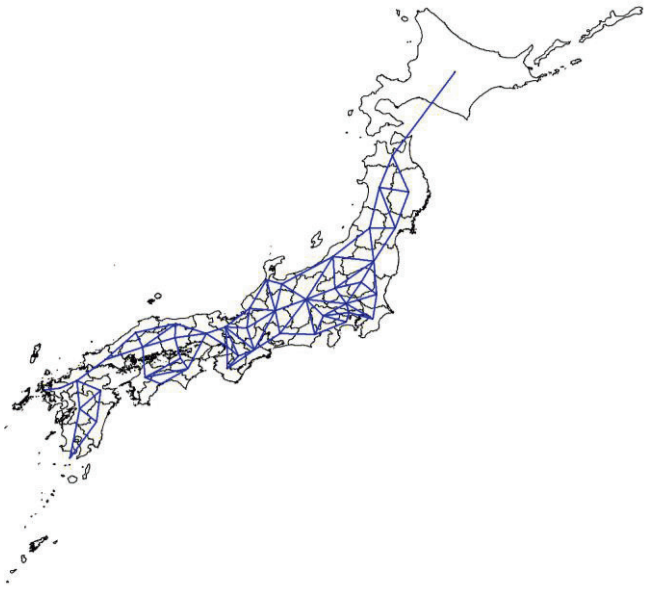

図５　都道府県間の隣接関係のグラフ

4.2 推定手法

本稿の時空間モデルは，時間依存，空間依存，時空間依存のランダム効果や回帰係数それぞれに事前分布が導入され，さらにその事前分布の超パラメータにも超事前分布が与えられた階層ベイズモデルとなっている．階層ベイズモデルにおけるパラメータ推定は，推定すべきパラメータの事後分布，すなわち，データが与えられた下でのパラメータの条件付き分布を得た上で，事後分布の平均を推定値（事後平均と呼ぶ）とし，事後分布の下側 2.5%点から上側 2.5%点までの区間を 95%信頼区間（95%ベイズ信頼区間と呼ぶ）とするのが一般的である．

パラメータの事後分布を求めるための代表的な手法として，乱数を用いた一定の規則によるパラメータ更新を繰り返すことでパラメータの事後分布に近似的に従うサンプルを得るマルコフ連鎖モンテカルロ法がある．本研究でも当初は R の Nimble パッケージや CARBayesST パッケージを用いたマルコフ連鎖モンテカルロ法による推定を検討したものの，モデルの複雑さとデータ量ゆえに相当の更新回数を経てもサンプルの従う分布が事後分布へ収束することを確認でき

なかった．一方，R の RStan パッケージでは事後分布への収束が速いハミルトニアンモンテカルロ法が実装されており，本分析にも利用できる可能性がある．ただし RStan パッケージでは一部の CAR モデルを扱う関数は存在するが，本分析で使用した Leroux モデルは実装されていないことから，Leroux モデルを扱う関数をもつ次に紹介する INLA パッケージを利用することとした．一方，佐野[2022]は保険データの時空間ベイズモデル解析において，マルコフ連鎖モンテカルロ法以外の選択肢として，サンプリングを用いない近似手法である INLA(Integrated Nested Laplace Approximation)を用いた解析例を示している．幅広い統計モデルを計算効率よく推定することのできる近似ベイズ推定論の INLA は，マルコフ連鎖モンテカルロ法のようにサンプリングを繰り返すことで事後分布を生成する代わりに，事後分布の密度を少数の点で評価したうえで，事後分布全体を近似する方法である．そこで本稿におけるパラメータの推定手法として INLA を採用し，前節の各モデル候補について R の INLA パッケージを用いて推定を行った．R の INLA パッケージでは，モデルの推定だけでなく，モデルの当てはまりの良さや予測精度を測るための指標も得ることができる．本研究では，INLA パッケージから得られるモデル選択基準として DIC (Deviance Information Criterion)(Spiegelhalter et al.[2002])，および WAIC (Widely Applicable Information Criterion)(Watanabe and Opper[2010]) を用いてモデル候補からの選択を行った．さらに，予測精度の指標として，対数周辺予測尤度(Log Marginal Predictive Likelihood: LMPL)，ならびに経過台数を重みとした加重平方平均二乗誤差 (Weighted Root Mean Squared Error: WRMSE) も算出して併せてモデル選択に用いた．

5 分析結果

5.1 モデル選択

表１はモデルごとの DIC，WAIC，LMPL および WRMSE を示している．各統計量は，値が小さいほど当てはまりまたは予測精度が良く (LMPL は 0 に近いほど良い)，今回検証した 8 つのモデルでは Model2，Model6 が高い予測精度を示していることがわかる．時間及び都道府県による変化を固定効果として導入した GLM に比べても高い予測精度を示している．Model6 における走行距離指数の係数$\beta_{mileage}$の事後平均は 0.695 を示しており，95%ベイズ信頼区間も (0.213,1.107) と有意に正の値を取るため走行距離を加味した Model6 を採用する．なお，Model6 の方が交通量の変化に伴うクレーム頻度の変化をより適切に予測できることを，後の 5.5 節にて検証する．

また，INLA は高速な近似推定を可能としており，今回の分析における交互作用を考慮した複雑なベイズ時空間モデルであっても計算実行時間は約 2.5 秒ほどであり，短時間で計算を行うことができた．

	DIC	WAIC	LMPL	WRMSE
GLM	6041.9	117574.6	-13622.7	1442.8
Model1	6391.0	7605.3	-3355.5	341.1
Model2	5264.6	5137.2	-2890.5	6.1
Model3	6039.9	117616.9	-13622.0	1442.8
Model4	5283.3	5161.4	-2960.0	8.3
Model5	6418.0	7414.0	-3362.7	336.7
Model6	**5265.1**	**5138.0**	**-2892.3**	**6.0**
Model7	8535.0	30428.5	-6044.2	809.0
Model8	5273.4	5149.9	-2924.1	7.9

表１　各モデルの当てはまりの良さの指標比較

5.2 推定結果：空間依存のランダム効果

Model6 における Leroux モデルのρの推定値（事後平均）は 0.642 を示しており，95%ベイズ信頼区間も(0.272, 0.914）と有意に正の値を取ることからクレーム頻度のモデルにおいて正の空間相関が確認できた．なお，$\log\frac{\rho}{1-\rho}$に対して INLA で設定されている事前分布$\mathrm{N}(0, 0.45^{-1})$の精度 0.45 を 2 倍および半分へと変更した場合，ρの推定値はそれぞれ約 0.03 の変化があり，若干ながら事前分布の設定により影響を受けているものの，DIC， WAIC は高々0.02 の変化しか生じなかったため，事前分布の設定は当てはまりの良さに大きな影響を与えてはいないものと考えられる．

なお，図６および表２に示す空間依存のランダム効果u_kの事後平均すなわち推定値を比較すると，最も大きい値を示したのは福岡県の 0.332（指数変換すると 1.393）であり，最も小さい値を示した岩手県では-0.503（0.605）となり都道府県間で最大 2.3 倍のクレーム頻度の較差をもたらしている．

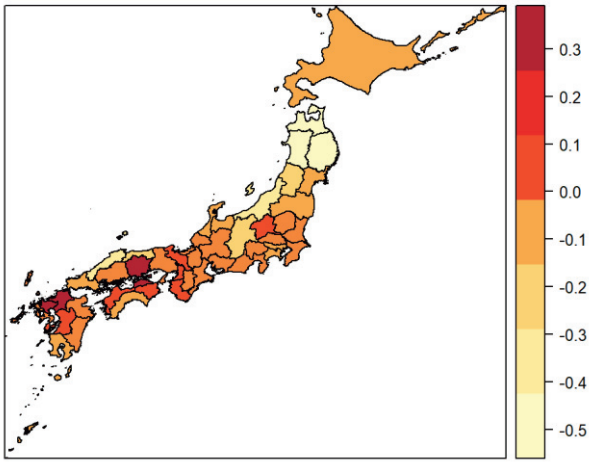

図６　時空間モデルの空間依存のランダム効果の推定値（事後平均）

また，空間依存のランダム効果u_kにおける推定値の地域的傾向を確認すると大きい値を示している福岡県の周辺の佐賀県や熊本県，香川県の周辺の徳島県，愛媛県，岡山県などは同様に高い値を示している傾向があり，青森県，岩手県，秋田県といった事後分布の平均の小さい地域も隣接している都道府県において同様に低い値を示していることがわかる．一方でu_kの値が大きい岡山県に隣接している鳥取県や群馬県と隣接している新潟県，長野県などは低い値を示しており，必ずしもすべての都道府県の隣接において正の空間相関が存在するわけではないことが確認できた．岡山県と鳥取県の間で相関が弱い理由としては，両県が中国山地により隔たれているため，表３に示すように互いの行き来が少ないことが考えられる．新潟県と群馬県の間も同様であり，長野県に関しては群馬県以外にも７つの県と隣接して行き来があるために群馬県との相関が強くはならなかったものと考えられる．

順位	都道府県	u_k	$\exp(u_k)$	順位	都道府県	u_k	$\exp(u_k)$
1	福岡県	0.332	1.393	24	三重県	0.048	1.049
2	香川県	0.325	1.384	25	埼玉県	0.033	1.034
3	岡山県	0.269	1.309	26	広島県	0.028	1.028
4	佐賀県	0.268	1.308	27	大分県	0.026	1.026
5	群馬県	0.215	1.240	28	長崎県	0.013	1.013
6	徳島県	0.204	1.227	29	石川県	-0.040	0.961
7	熊本県	0.189	1.208	30	山口県	-0.052	0.950
8	大阪府	0.181	1.198	31	神奈川県	-0.057	0.944
9	愛媛県	0.141	1.151	32	富山県	-0.058	0.944
10	京都府	0.131	1.140	33	東京都	-0.058	0.944
11	和歌山県	0.121	1.129	34	宮城県	-0.062	0.940
12	奈良県	0.113	1.119	35	福島県	-0.100	0.905
13	岐阜県	0.112	1.118	36	鹿児島県	-0.105	0.900
14	兵庫県	0.109	1.115	37	北海道	-0.109	0.896
15	静岡県	0.100	1.105	38	高知県	-0.126	0.882
16	福井県	0.092	1.096	39	鳥取県	-0.212	0.809
17	愛知県	0.085	1.089	40	長野県	-0.238	0.788
18	茨城県	0.085	1.089	41	山形県	-0.252	0.778
19	山梨県	0.071	1.074	42	新潟県	-0.304	0.738
20	宮崎県	0.070	1.073	43	島根県	-0.327	0.721
21	栃木県	0.069	1.072	44	青森県	-0.458	0.633
22	滋賀県	0.067	1.069	45	秋田県	-0.493	0.611
23	千葉県	0.058	1.060	46	岩手県	-0.503	0.605

表２　都道府県ごとの空間依存のランダム効果u_kの推定値（事後平均）

出発地	目的地	トリップ数／日
群馬県	福島県	280
	栃木県	50,240
	埼玉県	69,741
	新潟県	**1,488**
	長野県	**5,154**
新潟県	山形県	2,034
	福島県	1,716
	群馬県	**1,687**
	富山県	1,469
	長野県	4,952
長野県	**群馬県**	**4,937**
	埼玉県	3,061
	新潟県	4,917
	富山県	661
	山梨県	5,601
	岐阜県	4,289
	静岡県	1,266
	愛知県	5,303
岡山県	兵庫県	7,682
	鳥取県	**3,972**
	広島県	31,787
	香川県	1,411
鳥取県	兵庫県	3,224
	島根県	21,712
	岡山県	**3,658**
	広島県	1,236

表３　平成 27 年度全国道路・街路交通情勢調査自動車起終点調査より
乗用自動車における目的都道府県別トリップ数の比較

5.3 推定結果：時間依存のランダム効果

図７は時空間モデルの時間依存のランダム効果γ_tの事後平均による推定値と，95%ベイズ信頼区間を示している．図２で見た年度別のクレーム頻度の推移と同様に年度ごとの減少傾向が確認できる．

2021 年度の推定値は−0.210 と求まり，式(6)に従って 2021 年度のトレンドを引き延ばして得た 2022 年度の予測値は−0.264 という結果になった．これは指数変換するとそれぞれ 0.811 と 0.768 になり 2021 年度から 2022 年度にかけてクレーム頻度に対して約 4.3%減の影響を与える予測となっている．

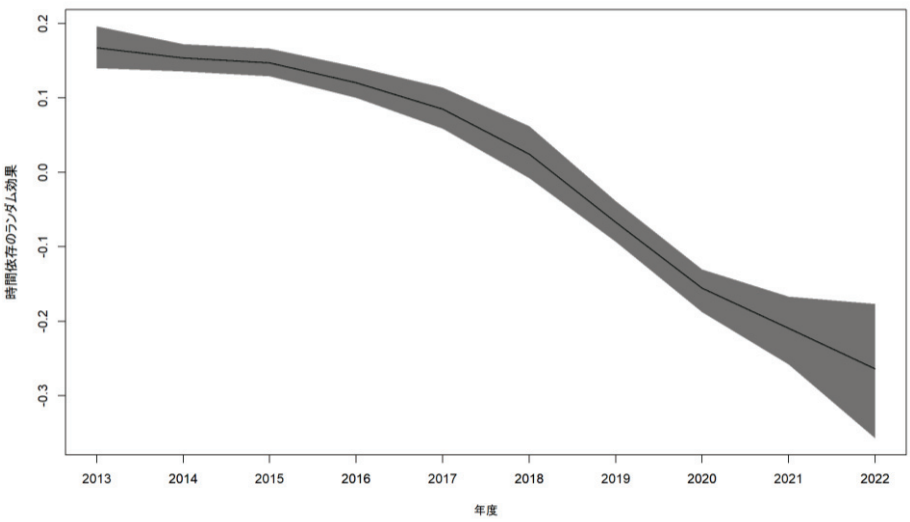

図７　時空間モデルの時間依存のランダム効果の事後分布の推定値（事後平均）と 95%ベイズ信頼区間

5.4　推定結果：時空間依存のランダム効果（時空間の交互作用）

時空間の交互作用については独立同一分布を仮定していたが，図８に示した推定値にはやや系統的な時空間変動が見て取れる．各地域や年度ごとに特徴がみられ，北海道，群馬県，神奈川県，岐阜県，三重県，大阪府，兵庫県，徳島県，福岡県，大分県といった地域は上昇傾向にある一方で，岩手県，宮城県，山形県，福島県，富山県，福井県，山梨県，高知県，佐賀県，宮崎県は減少傾向にあることがわかる．

全国的に事故頻度は減少傾向にあるものの，減少の速度には都道府県間に較差があり，すべての都道府県で一様に減少しているわけではないことが交互作用によって確認された．

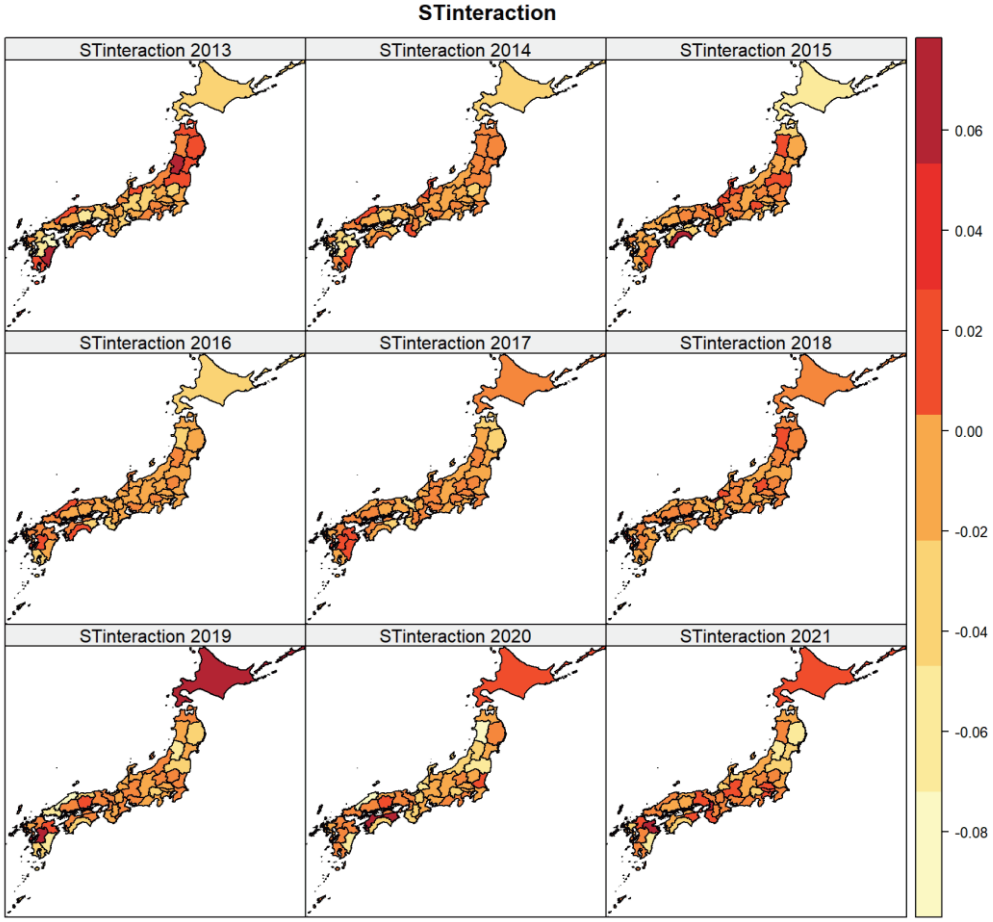

図８　時空間モデルの交互作用の推定値（事後平均）

5.5　2022 年度のクレーム頻度予測

Model6 を採用し，Model6 から得られたランダム効果および 2022 年度の走行距離を基に 2022 年度のクレーム頻度の予測を行った結果を図９に示す.

2022 年度における走行距離は前年度に比べ全国平均で 9.1%増加しているが，時間のランダム効果による減少トレンドと相殺され，22 年度における全国のクレーム頻度の予測値は 1.008%となり前年度の 1.012%と比べほぼ横ばいあるいは微減という結果となった.

また図 10 には Model6 及び走行距離を考慮していない Model2 から得られたランダム効果を基に予測した 46 都道府県全体でのクレーム頻度のグラフを示している．Model2 では時間依存によるランダム効果による減少トレンドによって，2022 年度の予測値においても前年度と比べ減少している結果となっている.

これらの予測を実施した後，2023 年 9 月 29 日に 2022 年度の「自賠責保険統計」が公開された．そこから算出される 2022 年度のクレーム頻度の実績値は 1.025%と前年度の 1.012%からほぼ横ばいあるいは微増となり，走行距離を加味せず減少トレンドのみを考慮した Model2 の予測は大きく外れた一方，Model6 の交通量回復によるクレーム頻度上昇も加味して横ばいとしたクレーム頻度の予測と非常に近い結果となった．この結果から，クレーム頻度予測にあたって，交通量を加味することの重要性が示されたものと考える.

図９　都道府県別クレーム頻度

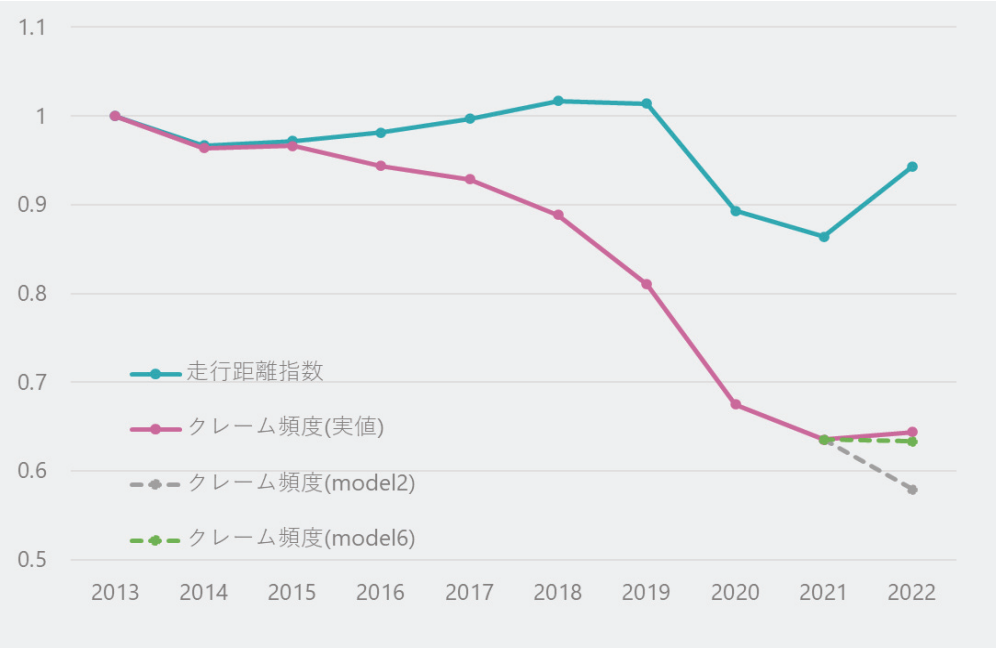

図 10　走行距離およびクレーム頻度の指数（2013 年度比）の推移

6 まとめ

本稿では，都道府県・年度ごとに集計された自家用乗用自動車の自動車損害賠償責任保険の支払件数データに対して，地域間の従属性および，新型コロナウイルスによる社会生活の変容の影響(自動車走行距離の変動)を加味したベイズ時空間モデルを用いて分析を行い，各都道府県のクレーム頻度の推定ならびに予測を行った．

自動車損害賠償責任保険の純保険料率（基準料率）は車種・用途・地域の３つのファクターのみで構成されている．特に地域は本土・沖縄本島・本土離島・沖縄離島の４地域区分で保険料率に較差を設けているが，本稿では46都道府県ごとのクレーム頻度の比較を行い，都道府県間の較差を明らかにした．一般にクレーム頻度の分析を行う手法としてGLMによるポアソン分布などを仮定した頻度モデルがあり，GLMの変数に地域を加えることで地域間の較差を反映することができる．最も単純な方法として都道府県や都道府県をグループ化したものをカテゴリ変数として組み込むことであるが，地域あたりのデータ量の減少や，データが地域間で分断されることによって信頼性が低下する恐れがある．そこで隣接する地域間の特徴を捉えることのできる手法として条件付き自己回帰モデルを採用した．

また自動車事故の発生件数は安全運転装置の普及もあり年々減少傾向にあること，新型コロナウイルスによる交通量の減少に伴った自動車事故の減少などを反映させるため，時間依存のランダム効果および共変量として「自動車燃料消費量調査」より取得した走行距離を組み込んだモデルを提案した．分析の結果，今回の各都道府県別のクレーム頻度の予測において隣接する都道府県において正の空間相関が確認でき，アフターコロナにおける走行距離の増加が事故頻度に影響を及ぼしていることが示された．さらに，2022年度の都道府県別クレーム頻度を予測した結果，従来から続くクレーム頻度の減少トレンドと，交通量の回復によるクレーム頻度の増加が打ち消しあい，2021 年度のクレーム頻度から概ね横ばいの予測となった．

なお，今回の提案モデルにおける予測では交互作用を考慮しておらず，交互作用の推定結果によって示された事故頻度の減少の速度には都道府県間に較差があるといった都道府県間のトレンドの違いを予測に組み込むことで，各都道府県の個別の変化も予測に反映してより高度な将来予測に発展することが考えられる．

また，今回の分析では空間依存のランダム効果における都道府県間の隣接関係$w_{k,j}$に対して地域kに対し地域jが隣接する場合は$w_{k,j} = 1$，そうでない場合は$w_{k,j} = 0$を設定しているが，空間依存のランダム効果の推定の結果，隣接しているにも関わらず空間相関の弱い都道府県が存在することが確認できた．国土交通省の公表している全国道路・街路交通情勢調査自動車起終点調査では各都道府県を出発地とした目的都道府県別のトリップ数の把握ができ，隣接していない都道府県へのトリップ数も把握することができる．そこで都道府県間の隣接関係の定義をトリップ数をもとに新たに設定することでより交通量の実態に則った時空間モデルへと拡張できると考えられ，今後の課題としたい．

謝辞

本研究にあたり，損害保険料率算出機構より損害保険料率算出機構統計集のデータの提供を受けた．また本研究の一部は，野村俊一へのROIS-DS-JOINT(025RP2023)の助成を受けて行われた．また，本論文の執筆にあたり，有益な助言を下さった査読者の方々に心より感謝する．

参考文献

[1] 佐野誠一郎(2022),「INLA による時空間の従属性を考慮した頻度モデル」.『会報(日本アクチュアリー会)』,Vol.75,105-130 頁.

[2] 島田亮(2021),「新型コロナウイルス感染症の感染拡大に伴う移動制限下での交通事故の発生件数, 負傷者数および死亡者数の変化」.『IATSS Review (国際交通安全学会誌)』,Vol.46,No.1,16-21 頁.

[3] 村上大輔(2022),『R ではじめる地理空間データの統計解析入門』,講談社

[4] 釜賀一輝, 秋山孝正(2022),「新型コロナの生活様式変化に起因する都市交通問題に関する考察」.『日本都市計画学会関西支部研究発表会講演概要集 20』,29-32 頁.

[5] 兵頭知(2022),「COVID-19 感染拡大に伴う交通状態変化が都市内高速道路における交通事故リスクに与えた影響に関する研究」.『交通工学研究発表会論文集 第 42 回交通工学研究発表会』. 一般社団法人 交通工学研究会,403-410 頁

[6] Assuncao R., M.A. Costa, M.O. Prates, and L.s.G. Silva e Silva(2014), Spatial analysis. In *Charpentier, A., editor, Computational Actuarial Science with R*. Boca Raton, FL: Chapman & Hall. P. pp.207-256.

[7] Besag, J.(1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society: Series B (Methodological),* Vol.36, No.2, pp.192-225.

[8] Besag, J., J. York, and A. Mollie(1991), "Bayesian Image Restoration, with Two Applications in Spatial Statistics," *Annals of the Institute of Statistical Mathematics,* Vol.43, No.1, pp.1-20.

[9] Boskov M. and R.J. Verall(1994). Premium rating by geographic area using spatial models. *Journal of the International Actuarial Association,* Vol.24, No.1, pp.131-143.

[10] Gschlößl, S., and C. Czado(2007), "Spatial modelling of claim frequency and claim size in non-life insurance," *Scandinavian Actuarial Journal*, Vol.2007, No.3, pp.202-225.

[11] Lee, D.(2011),"A comparison of conditional autoregressive models used in Bayesian disease mapping, " *Spatial and Spatio Temporal Epidemiology,* Vol.2, No.2, pp.79-89.

[12] Leroux, B.G., X. Lei, and N.E. Breslow(2000), "Estimation of Disease Rates in Small Areas: A New Mixed Model for Spatial Dependence," In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*(pp.179-191), Springer, New York.

[13] Ozaki, V.A., S.K. Ghosh, B.K. Goodwin, and R. Shirota(2008). Spatio‐temporal modeling of agricultural yield data with an application to pricing crop insurance contracts. *American journal of agricultural economics*, Vol.90, No.4, pp.951-961.

[14] Rue, H., S. Martino, and N. Chopin(2009), "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Vol.71, No.2, pp.319-392.

[15] Stern, H.S. and N. Cressie(1999), "Inference for extremes in disease mapping," In *Disease Mapping and Risk Assessment for Public Health*,(pp.63-84), Wiley, Chichester.

[16] Spiegelhalter, D.J., N.G. Best, B.P. Carlin, and A. Van Der Linde(2002), "Bayesian Measures of Model Complexity and Fit (with Discussion)," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* Vol.64, No.4, pp.583-616.

[17] Tufvesson, O., J. Lindström, and E. Lindström(2019). Spatial statistical modelling of insurance risk: a spatial epidemiological approach to car insurance. *Scandinavian Actuarial Journal*, Vol.2019, No.6, pp.508-522.

[18] Watanabe, S., and M. Opper(2010), ”Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory,” *Journal of Machine Learning Research,* Vol.11, No.12.

Statistical analysis of car accidents frequency by prefecture during the COVID-19 pandemic with consideration of traffic volume fluctuations

Shota Tanikawa* Shunichi Nomura† Seiichiro Sano‡

Abstract

The COVID-19 pandemic not only inflicted severe damage on human lives but also ushered in significant changes in societal lifestyles. Notably, travel restrictions imposed by state-of-emergency declarations and the proliferation of teleworking for infection prevention led to a drastic reduction in traffic volumes, which noticeably reduced the number of traffic accidents. However, as vaccination rollout progressed and the virus mutations led to a reduction in the fatality rate of the infection, these behavioral restrictions began to ease. Consequently, traffic volumes have been gradually rebounding to their previous levels. From the perspective of non-life insurance companies, it is vital to accurately predict the fluctuations in claim frequency in line with the resurgence in traffic for insurance pricing and financial management purposes.

In this study, we employed a spatiotemporal Bayesian model to analyze the claim frequency of compulsory automobile liability insurance by prefecture from 2013 to 2021, as disclosed by General Insurance Rating Organization of Japan. By introducing the vehicle travel distance as a covariate, we discuss the impact of the reduction in traffic volumes on claim frequency. Furthermore, we forecast the claim frequency by prefecture for the fiscal year 2022.

Keywords： COVID-19, Compulsory automobile liability insurance, Claim-frequency analysis, Spatiotemporal Bayesian model

* Faculty of Commerce, Graduate School of Accountancy, Waseda University Email: shota.tanikawa@toki.waseda.jp
† Faculty of Commerce, Graduate School of Accountancy, Waseda University Email: snomura5@aoni.waseda.jp
‡ Kyoei Fire & Marine Insurance Co.,Ltd. Email: se.sano@kyoeikasai.co.jp