

冬期高速道路における交通事故の発生要因分析とリスク評価

学籍番号：19108187 都市交通研究室 渡邊 和志

指導教員 佐野 可寸志

1. 研究の背景と目的

冬期に発生する交通事故は、大雪や吹雪、強風、といった気象条件の悪化によって引き起こされる。これらの気象要因は、ドライバーの視界不良や路面の凍結や積雪による路面状態の悪化をもたらし、安全な運転を阻害する主な要因となっている。東日本高速道路株式会社から提供していただいた交通事故データを基に分析を行った結果、図 1-1 に示すように、2016 年 1 月から 2020 年 12 月の新潟県における高速道路の本線内で発生した月ごとの交通事故件数は、他の季節と比較して 12 月、1 月、2 月に特に多いことが分かった。中でも 1 月では最多の 1069 件となっており、冬期特有の道路条件や気象条件が交通事故発生件数の増加に影響を与えていると考えられる。

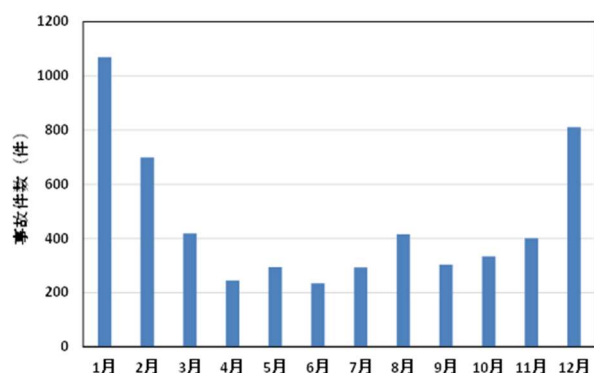


図 1-1 2016 年から 2020 年における新潟県内の高速道路での月別交通事故発生件数

これらの背景を踏まえ、本研究では、新潟県を通る高速道路を対象に、冬期の高速道路における交通事故発生に関わる要因を分析し、交通事故の発生リスクを推定するモデルを構築することを目的とする。具体的には、交通状況、気象条件、道路線形、路面状態といった要因を考慮し交通事故リスクの高い区間や時間帯を定量的に評価する。このモデルを活用することで、事故の発生リスクが高い時間帯や区間の計画的通行止めや事故の発生しやすい区間の整備・改善策の検討に活用することを目指す。

これにより、冬期における高速道路の安全性向上と交通事故の減少に貢献する。

2. 既往研究の整理

本研究は冬期の高速道路における交通事故の発生を対象とした研究である。本研究を行うにあたり関連する既往研究について以下に整理する。

渡部ら¹⁾は、交通事故削減に向けた効果的・効率的な対策実施を目標として、事故情報に道路構造、交通状況、土地利用状況などを結合した事故分析データベースを GIS 上で構築し、道路交通環境や社会環境条件と事故発生要因との関係を負の二項回帰分析と重回帰分析を用いて事故類型別に分析を行った。その結果から、幹線道路の事故には旅行速度や適切な交差点距離の設定といった、走行円滑性の向上に繋がる要因が安全性の向上に寄与していることを明らかにした。また、交通情報の少ない非幹線道路については、歩行者や自転車利用者の交通量を代替すると考えられる指標が歩行者や自転車の事故件数に統計的有意に作用していることや、幅員別道路延長量や用途地域などの要因では、事故発生件数に最も大きく影響するカテゴリーが事故類型間で異なることを明らかにした。

Poch ら²⁾は、交差点における交通事故削減に向けた効果的な対策の検討を目的として、ワシントン州ベルビュー市の 63 の交差点における 7 年間の事故データに対して、負の二項回帰モデルを用いた事故発生確率の分析を行った。その結果から、交差点における事故発生確率は、交通量や左折・右折の交通量、信号制御の有無、視程制限の有無といった要因の影響を受けることを明らかにした。特に、左折交通量の増加が事故発生に対して最も強い影響を与え、左折専用レーンの設置や信号制御の適正化が事故削減に有効であることが示唆された。また、視距制限のある交差点では事故発生リスクが統計的に有意に増加することが確認され、適切な視距の確保が安全対策として重要であることが明らかにした。

佐野ら³⁾は、寒冷・積雪地域における中央分離帯のない高速道路の交通事故減少のための効果的な対策の提案を目標として、新潟県内における中央分離帯がない高速道路に対して、発生した交通事故を分析した。交通事故を死亡・重傷と軽傷の二つに分類し、判別分析法を用いて事故の重大性や事故リスクに影響を与える要因の分析を行った。分析結果より、中央分離帯がない高速道路での事故の特徴として、中央分離帯の

ある高速道路に比べて死亡事故率が高いこと、特に凍結路面での死亡事故率が高いことを明らかにした。また、速度の増加、急勾配、右カーブは重大な事故の発生に大きな影響を与える要因であることを明らかにした。

Manh ら⁴⁾は、冬期の積雪条件下での交通事故の発生要因を明らかにし、1時間単位での交通事故発生確率を推定するモデルの構築を目的とした。分析方法として、新潟県内の高速道路で発生した交通事故を対象として事故データセットを作成し、負の二項回帰モデルを用いて1時間当たりの交通事故発生に与える要因の分析を行い、交通事故発生確率を推定した。分析結果より、交通量や降雪量が交通事故発生確率に正の影響を与える一方、大型車混入率や気温、平均車速が事故発生確率を減少させることが明らかとなった。しかし、データにゼロ過剰が含まれる可能性や路面状態や積雪深の影響を十分に考慮できていない点が課題として挙げられている。

四辻ら⁵⁾は、北海道を通る道央自動車道、札幌自動車道、国道36号を対象とし、高速道路と一般道路が並行する区間に対して、冬期の交通事故リスク要因の回帰分析を行い、並行ルート間の冬期の交通事故リスクと交通量転換との関係についての分析を行った。また、モデルの選択において過分散検定とゼロ過剰検定を行うことで統計的に最適なモデルを採用した。分析結果より、一般道路の事故リスクは信号交差点密度や降雪量、積雪深などと相関し、高速道路の事故リスクはキロ当たり交通量や降雪量、発生月などと相関することを明らかにした。また、冬期交通事故リスク情報を基にした交通量転換シミュレーションにより、一般道路から高速道路への交通量転換が特定の条件下で事故リスクの低減に寄与することを示した。

以上の研究から、交通事故発生に影響を与える要因についての分析は多く行われているが、冬期の高速道路で発生した交通事故に関して、路面状態を考慮した短時間における交通事故発生要因に関する分析を行ったものはほとんどみられない。

そこで本研究では、新潟県を通る高速道路を対象に、1時間単位かつIC間単位での交通事故データセットを作成した。さらに、作成したデータセットに対して、複数の一般化線形モデルを適用し、先行研究と同様に過分散検定とゼロ過剰検定を行うことで、データ特性にあった適切なモデルを選択した。

3. 研究の全体像

3.1 使用データの概要

事故データセットは、新潟県を通る日本海東北自動車道、北陸自動車道、磐越自動車道、関越自動車道、

上信越自動車道（図3-1）を分析の対象区間として、2016年1月から2020年の12月の冬期間（12月、1月、2月、3月）を対象期間として作成した。また、交通事故の発生要因に関する分析と交通事故リスクの推定には、同様の事故データセットを用いた。

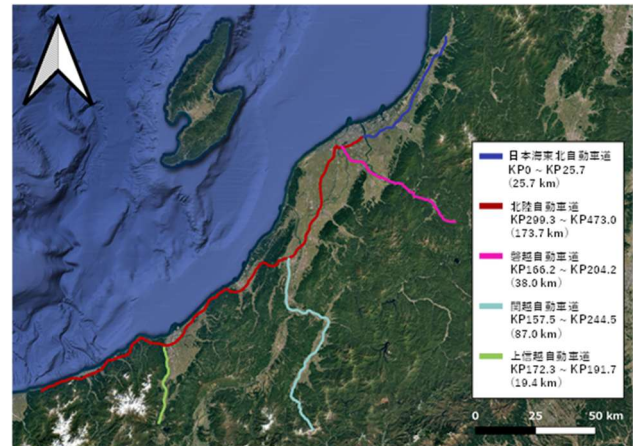


図 3-1 分析対象の高速道路区間

事故データセットの作成に使用したデータの概要を表3-1に示す。事故データセットには、交通事故データ、交通状況データ、気象データ、道路線形データ、路面状態の推定データが含まれる。

表 3-1 事故データセットの概要

使用データ	変数
交通事故データ	事故件数(件)
	交通量(台/h)
交通状況データ	区間長(km)
	平均車速(km/h)
	大型車混入率(%)
	平均気温(°C)
気象データ	最低気温(°C)
	1時間降雪量(cm/h)
	3時間降雪量(cm/3h)
	6時間降雪量(cm/6h)
道路線形データ	最大上り勾配(%)
	最大下り勾配(%)
路面状態の推定データ	路面の推定結果

本研究で使用する交通事故データ、交通状況データ、気象データ、道路線形データは、東日本高速道路株式会社から提供していただいたデータを用いた。また、路面状態の推定データは、戸澤ら⁶⁾が作成した路面推定モデルにより推定した路面状態を用いた。それぞれのデータについては、以下の手順で集計を行った。

ここで、事故データセットは、データごとに算出し

た変数を1時間単位かつIC間単位で集計した。1時間単位での集計を行った理由としては、降雪量や路面状態が交通事故発生に影響をより明確に把握するためである。また、IC間単位での集計を行った理由としては、実務でモデルを活用する際に通行止め区間に適合させるためである。

a) 交通状況データ

対象区間における交通量は、対象区間内の車両感知器で観測された交通量の合計を、対象区間に含まれる車両感知器の数で除した値として算出した。対象区間における平均車速は交通量と同様に、対象区間内の車両感知器で観測された平均車速の合計を、対象区間に含まれる車両感知器の数で除した値として算出した。対象区間における大型車混入率は、対象区間に含まれる大型車の台数を対象区間に含まれる全車両の交通量で除した値として算出した。

また、対象区間に対応する車両感知器に欠損データがある場合には、次に近い車両感知器のデータを用いた。この場合の近さは、対象区間と車両感知器の高速道路上の距離を基準として判断した。

b) 気象データ

対象区間における降雪量は、対象区間に最も近い観測所で観測された降雪量を1時間あたりの合計値、3時間あたりの合計値、6時間あたりの合計値を算出した。最低気温と平均気温は、対象区間に最も近い観測所で観測された気温の1時間あたりの最低値と平均値を算出した。

また、対象区間に対応する観測所に欠損データがある場合には、次に近い観測所で観測されたデータを用いた。この場合の近さは、対象区間と各観測所の緯度経度を基に計算した直線距離を基準として判断した。

c) 道路線形データ

対象区間における道路線形は、対象区間内に含まれる観測された縦断上り勾配の最大値を最大上り勾配、縦断下り勾配の最大値を最大下り勾配として算出した。

d) 路面状態の推定データ

対象区間における路面状態は、戸澤らが構築した路面状態の推定モデルを用いた。この推定モデルは、東日本高速道路株式会社のパトロールカーの車上からスマートフォン等で撮影した路面画像を基に、目視判別によって作成された教師データを用いている。教師データには、乾燥、濡れ、水たまり、冠水、湿雪、乾雪、圧雪、凍結といった、道路管理に合わせた路面状

態の分類が含まれている。この教師データを用いて機械学習手法の一つであるLightGBMを用いて路面状態の推定モデルを構築した。

本研究では、対象区間ごとの路面状態を推定するために、交通状況データ、気象データ、道路線形データを1時間単位かつIC間単位で集計した。その後、1時間降雪量、平均気温、平均車速といった変数を用いて、LightGBMを用いた推定モデルによって1時間単位かつIC間単位における路面状態を推定した。表3-2に、路面状態の推定結果を示す。

表 3-2 路面状態の推定結果

路面状態	事故なし(件)	事故あり(件)	事故件数の割合
乾燥	60,854	334	0.546%
濡れ	1,111,185	1,594	0.143%
薄雪	3,321	30	0.895%
湿雪	55,730	295	0.527%
凍結	1,775	24	1.334%
乾雪	11,557	154	1.315%
圧雪	1,077	12	1.102%

表3-2に示されるように、路面状態の推定結果についてまとめると、濡れ路面が推定結果のサンプル数の大半を占めている。しかし、その事故件数の割合は、他の路面状態と比較して低い結果となった。一方で、乾雪路面や薄雪路面、凍結路面、圧雪路面はサンプル数が少ないにもかかわらず、事故件数の割合が高い傾向が見られた。

本研究では、サンプル数が少なく、事故件数の割合が他の路面と比較して高い、薄雪路面、凍結路面、圧雪路面の3つの路面状態をまとめ、薄雪・凍結・圧雪路面とした。

3.2 使用モデル

(1) 交通事故発生要因を分析するモデルの構築

交通事故の発生に影響を及ぼす要因に関する分析では、一般に交通事故分析で用いられるポアソン回帰モデル、負の二項回帰モデルに加えて、ゼロ過剰データに対応したゼロ過剰ポアソン回帰モデル、ゼロ過剰負の二項回帰モデルを構築した。なお、モデルの構築にはプログラミングソフトRを用いた。

1) ポアソン回帰モデル

ポアソン回帰モデルは目的変数 Y が平均 μ 、分散 μ を分布パラメータにもつポアソン分布に従うと仮定すると、次式で定義される。

$$P[Y] = \frac{\mu^Y}{Y!} \exp(-\mu) \quad (1)$$

$$\mu = \exp(\alpha + \beta_1 X_1 + \dots + \beta_n X_n) \quad (2)$$

ここで、
 Y : 事故件数
 $P[Y]$: Y の発生確率
 μ : Y の期待値
 α : 切片
 β_n : 係数
 X_N : 説明変数

Y を事故件数とした場合、式(1)より事故が起きない確率 $P[Y = 0]$ と 1 件以上起きる確率 $P[Y > 0]$ はそれぞれ、 $P[Y = 0] = \exp(-\mu)$ 、 $P[Y > 0] = 1 - \exp(-\mu)$ となる。

2) 負の二項回帰モデル

負の二項回帰モデルは、ポアソン回帰モデルの拡張したものある²⁾。目的変数 Y が説明変数に依存する平均値 μ の負の二項分布に従うと仮定し、式(2)に $\log(\varepsilon)$ を加えることで過分散性を考慮すると、次式が成り立つ。

$$P[Y] = \frac{\Gamma(Y + \theta)}{Y! \cdot \Gamma(\theta)} \left\{ \frac{\theta}{\mu + \theta} \right\}^\theta \left\{ \frac{\mu}{\mu + \theta} \right\}^Y \quad (3)$$

$$\mu = \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \log(\varepsilon)) \quad (4)$$

ここで、
 Y : 事故件数
 $P[Y]$: Y の発生確率
 Γ : ガンマ関数
 θ : サイズパラメータ
 μ : Y の期待値
 β_n : 係数
 X_N : 説明変数
 ε : 平均 1, 分散 $1/\theta$ を分布パラメータにもつガンマ分布に従う誤差

Y を事故件数とした場合、式(3)より事故が起きない確率 $P[Y = 0]$ と 1 件以上起きる確率 $P[Y > 0]$ はそれぞれ、 $P[Y = 0] = \{\theta / (\mu + \theta)\}^\theta$ 、 $P[Y > 0] = 1 - \{\theta / (\mu + \theta)\}$ となる。

3) ゼロ過剰ポアソン回帰モデル

ゼロ過剰ポアソン回帰モデルでは、0 の値の起こりやすさと 1 以上の値の起こりやすさの比をロジスティック回帰モデルで推計し、次にそれぞれの起こりやすさをポアソン回帰モデルで推計する⁷⁾。目的変数 Y をゼロ過剰ポアソン分布に従うとすると、 Y の母集団から 0 件という標本が抽出される確率を π 、1 件以上という標本が抽出される確率を $1 - \pi$ とおく。式(1)よ

り、事故件数 0 件の確率は $\exp(-\mu)$ 、 k 件の確率は $(\mu^k / k!) \exp(-\mu)$ である。このとき、 $P[Y = k]$ は、 Y の母集団から 0 件という標本が抽出される確率、または、ポアソン分布で k 件となる確率の同時確率になる。

以上より、ゼロ過剰ポアソン回帰モデルは、次式で定義される。式(5)がポアソン回帰部分を、式(6)がロジスティック回帰部分を表している。

$$\mu = \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_n X_n) \quad (5)$$

$$\frac{\pi}{1 - \pi} = \exp(\gamma_0 + \gamma_1 X_1 + \dots + \gamma_n X_n) \quad (6)$$

$$P[Y] = \begin{cases} \pi + (1 - \pi) \exp(-\mu) & Y = 0 \\ (1 - \pi) \frac{\mu^Y}{Y!} & Y \geq 1 \end{cases} \quad (7)$$

ここで、
 Y : 事故件数
 μ : Y の期待値
 β_n : 係数
 X_N : 説明変数
 γ_n : 係数

$\hat{\pi}$ の推定に対して EM アルゴリズムに基づく最尤法を適用し、 $\hat{\beta}$ と $\hat{\gamma}$ を同時に推定する。また、ポアソン分布では平均 μ と分散 μ は等しくなるのに対し、ゼロ過剰ポアソン分布では平均 $(1 - \pi)\mu$ と分散 $(1 + \pi\mu)(1 - \pi)\mu$ は等しくならない。

4) ゼロ過剰負の二項回帰モデル

ゼロ過剰負の二項回帰モデルは、次式で定義される。ただし、式中の π と μ は、式(5)、(6)と同じく定義される。

$$P[Y] = \begin{cases} \pi + (1 - \pi) \left\{ \frac{\theta}{\mu + \theta} \right\}^\theta & Y = 0 \\ (1 - \pi) \frac{\Gamma(Y + \theta)}{Y! \cdot \Gamma(\theta)} \left\{ \frac{\theta}{\mu + \theta} \right\}^\theta \left\{ \frac{\mu}{\mu + \theta} \right\}^Y & Y \geq 1 \end{cases} \quad (8)$$

ここで、
 Y : 事故件数
 θ : サイズパラメータ
 μ : Y の期待値
 Γ : ガンマ関数

ここで、ゼロ過剰ポアソン回帰モデルと同様に、EM アルゴリズムに基づく最尤法を適用し、 $\hat{\theta}$ と $\hat{\beta}$ と $\hat{\gamma}$ を同時に推定する。オフセット項とは、一般化線形モデルや回帰モデルにおいて、説明変数として係数を持たず、

固定された値として扱われる項のことであり、分析対象のデータが異なる観測期間や母集団サイズを持つ場合に、オフセット項を導入することで、比較可能な確率を推定することができる⁸⁾。

本研究では、交通事故発生要因に関する分析および交通事故発生リスクを推定する際に、交通量と区間長をオフセット項として導入した。理由として、交通量が多い道路区間では、事故件数が増える傾向にあるため、オフセット項を導入することで、交通量あたりの事故発生確率を推定することができる。同様に、区間長が長い道路区間では、事故件数が増える傾向にあるため、区間長をオフセット項として導入することで、単位距離あたりの事故発生確率を推定することができる。また、これらの変数をオフセット項とすることで、気象条件や道路線形、路面状態などの変数の影響を明確に把握することができる。

(2) 交通事故リスクを推定するモデルの構築

交通事故リスクの推定には交通事故発生要因に関する分析にて選択されたモデルに加えて、機械学習により交通事故発生確率を推定する LightGBM を構築した。なお、本研究では1時間かつIC間あたりの交通事故発生確率を交通事故リスクと定義する。

この手法は、勾配ブースティングを用いて、決定木分析を複数回行い、分析精度を向上させた手法である。また、勾配ブースティングとは、予測値と実測値の誤差を計算し、誤差を決定木分析で学習する方法である。加えて、Leaf-wise の考えに従って、決定木が分割されることにより、短時間で高性能な推定を可能とする特性を有する。LightGBM は、勾配に基づく片側サンプリング (GOSS) と排他的特徴束ね (EFB) を用いて、他の手法よりも高速なトレーニングと推論を実現し、大規模なデータセットでもメモリ効率に優れる特性がある。また、少量のデータでも高精度な推定が可能であり、不均衡なデータセットに対しても効果的に対応したモデルとなっている。

4. 交通事故発生要因の分析

4.1 分析の流れとモデルの選択手法

本研究で分析対象としている交通事故は稀な事象であり、使用データ内の事故件数は表 4-1 に示すように、多くの観測値で交通事故が発生していないことが特徴として挙げられる。これにより、データに含まれるゼロの値の割合が非常に高いゼロ過剰データである場合や事故発生件数の分散が平均を上回る過分散である場合が考えられる。このようなデータ特性を持つ場合、前章にて説明したポアソン回帰モデルを用いると、ポアソン回帰モデルは目的変数がポアソン分

布に従うことを仮定するが、この分布は平均と分散が等しいことを前提としている。一方で、過分散が存在するデータでは、分散が平均を大きく上回るため、ポアソン回帰モデルでは誤差分布を適切に説明できず、モデルの信頼性が低下する。また、ゼロ過剰の特性を持つデータでは、事故が発生しない観測値の割合が異常に高いため、ポアソン回帰モデルではゼロの頻度を過小評価してしまう。そこで、ポアソン回帰モデルに加えて、過分散データを適切に扱うことのできる負の二項回帰モデル、データにおけるゼロの割合が高い場合に用いられるゼロ過剰モデルを構築した。ゼロ過剰モデルはロジスティック回帰で 0 件の発生確率を推定し、1 件以上の部分をポアソン分布や負の二項分布で推定することで、過剰なゼロを考慮することができる。これらのデータ特性に対処するためには、過分散やゼロ過剰を考慮した適切な統計モデルを選択する必要がある。

本研究では、先行研究と同様に過分散検定とゼロ過剰検定を行い、ポアソン回帰モデル、負の二項回帰モデル、ゼロ過剰ポアソン回帰モデル、ゼロ過剰負の二項回帰モデルの中から適切なモデルを選択した。なお、これらの検定には、プログラミングソフト R を用いた。

表 4-1 路面状態の推定結果

事故件数(件)	0件	1件
サンプル数(個)	1,111,185	1,594

① 過分散検定

過分散検定では、dispersiontest 関数を用いてポアソン回帰モデルに対して、ポアソン回帰モデルと負の二項回帰モデルにおける、平均と分散の差が 0 であるという帰無仮説を検定した⁹⁾。ここで、帰無仮説が棄却された場合には負の二項回帰モデルを、採択された場合にはポアソン回帰モデルを採用することで、適切なモデルを選択した。

② ゼロ過剰検定

ゼロ過剰検定では、vounG 関数を用いてポアソン回帰モデルとゼロ過剰ポアソン回帰モデルまたは、負の二項回帰モデルとゼロ過剰負の二項回帰モデルの両モデルに対して、それぞれのモデル間に有意差が無いという帰無仮説を検定した¹⁰⁾¹¹⁾。ここで、帰無仮説が棄却された場合には、ゼロ過剰ポアソン回帰モデルまたはゼロ過剰負の二項回帰モデルを、採択された場合には負の二項回帰モデルまたはポアソン回帰モデルを採用することで、適切なモデルを選択した。

4.2 モデルの適応結果とパラメータの推定結果

交通事故発生に影響を及ぼす要因に関する分析では、路面状態の有無が交通事故発生に与える影響を評価するため、路面状態の変数を含まないモデルと含むモデルを構築した。

ポアソン回帰モデル、負の二項回帰モデル、ゼロ過剰ポアソン回帰モデル、ゼロ過剰負の二項回帰モデルの各モデルに対する過分散検定とゼロ過剰検定の結果を表 4-2、表 4-3 に示す。表中で、それぞれの検定結果の右側にある「***」は 0.1% 有意、「**」は 1% 有意、「*」は 5% 有意であることを意味している。

表 4-2 各検定の結果（路面状態なしモデル）

	ポアソン	負の二項	ゼロ過剰ポアソン	ゼロ過剰負の二項
AIC	32,402	32,218	32,074	31,991
過分散検定	4.37 *** [NB]		—	
ゼロ過剰検定	—		-7.08 *** [ZINB]	

表 4-3 各検定の結果（路面状態ありモデル）

	ポアソン	負の二項	ゼロ過剰ポアソン	ゼロ過剰負の二項
AIC	32,333	32,169	32,039	31,959
過分散検定	4.26 *** [NB]		—	
ゼロ過剰検定	—		-6.50 *** [ZINB]	

表 4-2 より、路面状態を変数に含まないモデルでは、過分散検定の結果より、ポアソン回帰モデルに対して、平均と分散の差がゼロであるという帰無仮説が棄却されたため、負の二項回帰モデルを採用した。その後、ゼロ過剰検定の結果より、負の二項回帰モデルとゼロ過剰負の二項回帰モデルの両モデルに対して、モデル間に有意な差がないという帰無仮説が棄却されたため、ゼロ過剰負の二項回帰モデルを採用した。

表 4-3 より、路面状態を含むモデルに対して、過分散検定とゼロ過剰検定を行った結果、路面状態を含まないモデルと同様に、ゼロ過剰負の二項回帰モデルを採用した。

次に、路面状態を含まないモデルと、路面状態を含むモデルについて、それぞれのパラメータの推定を行った。説明変数の選択に際して、相関分析、VIF、AIC（赤池情報量基準）等の値を考慮し、統計的に適切な変数として選定した。表 4-4 に、路面状態を含まないモデルと路面状態を含むモデルに対するパラメータの推定結果を示す。なお、パラメータの推定値は標準化回帰係数である。表中で、推定値の右側にある「***」は 0.1% 有意、「**」は 1% 有意、「*」は 5% 有意であることを意味している。

表 4-4 より、採用したゼロ過剰負の二項回帰モデルを用いたパラメータの推定結果について考察する。推定結果より、説明変数には、両モデルとも共通して平均車速、大型車混入率、最低気温、6 時間降雪量、

表 4-4 路面状態の推定結果

	ゼロ過剰負の二項回帰モデル (路面状態なし)		ゼロ過剰負の二項回帰モデル (路面状態あり)	
切片	-13.7	***	-13.7	***
平均車速(km/h)	-0.371	***	-0.359	***
大型車混入率(%)	0.168	***	0.159	***
最低気温(C)	-0.576	***	-0.545	***
6時間降雪量(cm/h)	5.62.E-02	***	6.51.E-02	***
下り勾配	-8.74.E-02	**	-8.00.E-02	*
乾燥	—		0.470	***
湿雪	—		—	
乾雪	—		—	
薄雪・凍結・圧雪	—		0.308	*
Log θ	-1.36	***	-1.31	***
AIC	31,991		31,959	

最大下り勾配が有意な変数として選択された。

また、路面状態を含むモデルでは、乾燥、薄雪・凍結・圧雪の路面状態が選択された。推定された路面状態のパラメータの符号条件に着目すると、乾燥路面、薄雪・凍結・圧雪路面は正の係数を示し、有意な変数となっていた。これは、これらの路面状態において事故の発生リスクが増加する傾向を示している。このような結果となった理由として、乾燥路面では、勾配の緩やかな道路と同様に、安心して走行できるため、ドライバーの注意力散漫や速度超過により、事故リスクが高まると考えられる。薄雪・凍結・圧雪路面では、路面状態の悪化によるスリップにより、事故リスクが高まると考えられる。

これらのことから、平均車速、大型車混入率、最低気温、6 時間降雪量、最大下り勾配、乾燥路面、薄雪・凍結・圧雪路面が事故発生に影響を及ぼす要因として有意であることが分かった。

5. 交通事故リスクの推定と評価

5.1 交通事故リスクの推定方法

交通事故リスクの推定には、交通事故発生要因分析にて選択されたゼロ過剰負の二項回帰モデルに加えて、機械学習手法の一つである LightGBM を用いて事故発生リスクを推定した。理由として、ゼロ過剰負の二項回帰モデルは、事故が発生しやすい要因を確率的に説明することができ、本研究で使用する事故データのようなゼロ過剰データに対応しており、希少な事象に対する確率推定が可能となる。LightGBM は、決定木ベースのモデルであり、変数間の非線形な関係を自動的に捉えることが可能なため、多くの説明変数がある場合、変数同士の複雑な関係を考慮することができる。

本分析では、2016 年～2020 年のデータを学習用データとテストデータに分割し、学習用データには 4 年分、テストデータには 1 年分のデータを割り当てた。

このように異なる 5 パターンの学習用データとテストデータを用いることで、モデルの推定精度に対するデータセットの影響を排除した。それにより、特定の年に対して精度が高いモデルではなく、汎用性の高いモデルであるかを評価した。また、各学習用データに対してパラメータの推定を行い、事故発生リスクを推定した。

構築した事故発生リスクの推定モデルの精度を評価し、推定値と実際に事故が発生した区間との関係を定量的に評価するために、以下の手順での分析を行った。

まず、推定された事故発生リスクは 0~1 の範囲に収まるため、この範囲を 0.0001 刻みの階級値に分け、それぞれの推定値を該当する階級に分類した。これにより、細かいリスクごとでの詳細な分析が可能となる。次に、推定値に対応する実際に事故が発生した区間を分類した。その後、階級ごとに推定値の合計および実際に事故が発生した区間の合計を算出する。さらに、危険の基準とする階級値を設定し、階級値が基準値以上の場合は該当する区間を「危険」、基準値以下の場合は「安全」と分類した。

これにより、モデルごとの推定精度を比較し、事故発生リスクの推定精度の妥当性を評価する。

5.2 交通事故リスクの推定結果と考察

本研究では、交通事故リスクの推定結果を評価するために、危険と安全の区間のサンプル数の比率を 1:1, 1:5, 1:10 となるような基準値を設定した。理由として、交通事故リスクの分類において、モデルがどの程度事故の発生しやすい区間を識別できるかを評価するために、異なる基準値を用いて精度を比較する必要がある。危険区間を広げすぎると誤判定(偽陽性)が増え、逆に狭めすぎると見逃し(偽陰性)が増えるため、バランスの取れた基準値を見つけるためである。また、基準値を変化させることで、実際に事故が発生した区間の割合の変化を確認し、実務で活用する際に、危険とする基準をどこに設定するかが重要となるため、どの程度の基準値が最適であるかを検討することができる。

表 5-1 から表 5-2 に、ゼロ過剰負の二項回帰モデルと LightGBM を用いて、交通事故リスクを推定した結果を示す。推定結果より、両モデルとも、どのデータにおいても危険:安全の割合を 1:1, 1:5, 1:10 となるように基準値を変化させた際に同様の傾向がみられた。このことから、両モデルは特定の年に依存しない、汎用性のあるモデルであるといえる。

また、危険:安全の割合が 1:1 の場合は、「危険」と分類された区間に対して事故が発生しなかった区

間の割合が高く、偽陽性が増加している。1:5 の場合は、1:1 の場合と比較して「危険」と分類された区間の数が減少しているが、事故あり区間のうち「危険」と判断された割合は 1:1 よりも低下している。これは、基準値を厳しくすることで、より確実に事故が発生する区間を「危険」と分類しようとする一方で、事故が発生していた区間の一部を「安全」と誤分類してしまう可能性が高まるためと考えられる。しかし、偽陽性は 1:1 よりも減少している。1:10 の場合は、さらに「危険」と分類される区間の数が減少し、事故あり区間のうち「危険」と判断された割合も低下している。これは、より厳しい基準値を設定したことで、事故が発生する可能性が特に高い区間のみを「危険」と判断するようになったためである。しかし、実際に事故が発生した区間を「安全」と誤分類する割合が増加し、事故の見逃しリスクが高まる結果となっている。

以上の結果から、基準値の設定によって偽陽性と偽陰性のバランスが変化することが示唆される。1:1 では偽陽性が多くなり、1:10 では偽陰性が増加するため、実際の事故リスクを考慮しながら適切な基準値を設定することが重要であると考えられる。

表 5-1 事故発生リスクの推定結果
(ゼロ過剰負の二項回帰モデル)

a) 危険:安全が 1:1 (事故発生リスク=0.0012)

		実測値		総区間数
		事故あり	事故なし	
推定値	危険	401	111,748	112,149
	安全	73	138,264	138,337
計		474	250,012	250,486

b) 危険:安全が 1:5 (事故発生リスク=0.0024)

		実測値		総区間数
		事故あり	事故なし	
推定値	危険	289	41,180	41,469
	安全	185	208,832	209,017
計		474	250,012	250,486

c) 危険:安全が 1:10 (事故発生リスク=0.0038)

		実測値		総区間数
		事故あり	事故なし	
推定値	危険	240	22,853	23,093
	安全	234	227,159	227,393
計		474	250,012	250,486

表 5-2 事故発生リスクの推定結果 (LightGBM)

a) 危険 : 安全が 1 : 1 (事故発生リスク=0.0005)

		実測値		総区間数
		事故あり	事故なし	
推定値	危険	401	121,887	122,288
	安全	73	128,125	128,198
計		474	250,012	250,486

b) 危険 : 安全が 1 : 5 (事故発生リスク=0.0024)

		実測値		総区間数
		事故あり	事故なし	
推定値	危険	287	43,361	43,648
	安全	187	206,651	206,838
計		474	250,012	250,486

c) 危険 : 安全が 1 : 10 (事故発生リスク=0.0038)

		実測値		総区間数
		事故あり	事故なし	
推定値	危険	226	22,723	22,949
	安全	248	227,289	227,537
計		474	250,012	250,486

6. 結論

6.1 まとめ

本研究では、新潟県を通る 5 本の高速道路を対象に、冬期の高速道路における交通事故発生に影響を及ぼす要因の分析および交通事故の発生リスクを推定するモデルを構築した。使用データとして、交通事故データ、交通状況データ、気象データ、道路線形データ、路面状態の推定データをデータごとに 1 時間単位かつ IC 間単位での集計し、事故データセットを作成した。

はじめに、交通事故発生要因に関する分析では、ポアソン回帰モデル、負の二項回帰モデル、ゼロ過剰ポアソン回帰モデル、ゼロ過剰負の二項回帰モデルを適用し、過分散検定とゼロ過剰検定を行うことで、データの特性に適したモデルを選択した。各検定の結果、路面状態を含むモデルと含まないモデルの両モデルともゼロ過剰負の二項回帰モデルが適していることが確認された。各モデルでのパラメータの推定結果より、平均車速、大型車混入率、最低気温、6 時間降雪量、最大下り勾配、乾燥路面、薄雪・凍結・圧雪路面が事故発生に影響を及ぼす要因として有意であることが分かった。

次に、事故発生リスクの推定では、ゼロ過剰負の二項回帰モデルと LightGBM モデルを用いて推定を行った。推定の際に、異なる 5 パターンの学習用データとテストデータを用いることで、特定の年に対して精

度が高いモデルではなく、汎用性のあるモデルであるかを評価した。また、推定した事故発生リスクを階級分けし、階級ごとに推定区間の合計および実際に事故が発生した区間の合計を算出した。さらに、危険の基準とする階級値を設定し、階級値が基準値以上の区間を「危険」、基準値以下の区間を「安全」と分類し、危険と安全の区間の比率を 1 : 1, 1 : 5, 1 : 10 となるように基準値を設定し、モデルの評価を行った。推定結果より、両モデルとも、どのデータにおいても危険 : 安全の割合を 1 : 1, 1 : 5, 1 : 10 となるように基準値を変化させた際に、基準値が厳しい場合は、事故発生区間の的中率は向上したが、偽陰性が増加する傾向がみられた。一方で、基準値を緩和した場合は、事故の見逃しは減少したが、事故発生区間の的中率が低下し、偽陽性が増加するという傾向がみられた。このことから、両モデルは特定の年に依存しない、汎用性のあるモデルであることが確認できた。

実務における活用を考慮すると、両モデルとも同様な推定の傾向であったため、用途に応じて使用するモデルを決定する必要がある。機械学習モデルである LightGBM は、迅速なリスク評価をする際に有効であり、ゼロ過剰負の二項回帰モデルなどの回帰モデルは推定された係数を基に事故発生リスクに影響を与える要因を解析できるため、事故発生の要因を明らかにし、要因に応じて対策を検討する必要がある場合に有効であると考えられる。また、リスクの高い区間を重点的に管理するという目的の場合は、モデルの基準値を高く設定し、事故を見逃さないことを重要視する場合には、モデルの基準値を低く設定するなど、目的に応じて、適切な基準値を調整することが必要である。

6.2 今後の課題

本研究では冬期の高速道路における交通事故発生に影響を及ぼす要因の分析および交通事故の発生リスクを推定するモデルを構築した。結果として、事故発生に影響を与える要因を確認でき、事故発生リスクの推定モデルは、両モデルとも汎用性の高いモデルであることが確認できた。しかし、本研究が残している課題もいくつか存在する。それらの課題について以下に示す。

■ 交通事故の発生メカニズムについて

本研究では、交通事故発生に影響を及ぼす要因に関する分析を行ったが、パラメータの符号条件より、区間の平均車速が上昇すると、事故発生リスクは低下するという結果となった。また、路面状態に関して、乾燥路面、薄雪・凍結・圧雪路面である場合、事故発生リスクが増加するという結果となった。このような結

果になった理由として、事故データに含まれる事故の発生メカニズムが異なる事が考えられる。平均車速が低い事故の場合、渋滞や走行環境の悪化などによるものが考えられ、平均車速が高い区間の場合、乾燥路面での速度超過などが考えられる。これらの事故は発生メカニズムが異なるため、同じモデルで同時に分析するのではなく、事故の発生メカニズムごとにデータを分類し、モデルを構築する。または、分析対象とする事故以外の発生メカニズムによる事故をデータから除外することで、より精度の高い分析が可能になると考えられる。

■ 交通事故発生要因の分析および事故発生リスクの推定モデルの精度向上について

本研究では、交通事故データ、交通状況データ、気象データ、道路線形データ、路面状態の推定データを集計し事故データセットを作成し、事故発生要因の分析および事故発生リスクの推定を行ったが、ドライバー特性および車線数は考慮していないため、モデルの精度向上のために、ドライバー特性および車線数を考慮した分析を検討する必要があると考える。ドライバー特性を考慮することで、頻繁に速度超過を起こすドライバーは事故リスクが高まる可能性、長時間の運転をしているドライバーは注意力の低下により事故を起こしやすくなる可能性などを考慮することができる。また、車線数を考慮することで、追い越しによって発生する事故や、車線ごとの事故発生傾向を考慮したモデルの構築が可能となる。特に、降雪時には路面状態の悪化によって、走行可能な車線が限られる、または、走行車線に交通量が集中する可能性があるため、その影響を考慮することができる。そのため、モデルの精度向上のために、これらの変数の導入を検討する必要があると考える。

■ 構築したモデルの実務への活用について

本研究で対象とした高速道路には、すでに事故多発区間として、標識や警告表示、速度制限などの対策が講じられている区間が存在すると考えられる。よって、推定された事故発生リスクが等しい区間でも対策がされている区間とされていない区間とでは、事故リスクや検討する対策の内容が変わってくる。そのため、事故発生リスクの推定モデルの結果に併せて、実際の対策状況を考慮することで、既存の安全対策が十分に効果を発揮していない区間の特定や、新たな対策の検討をする必要があると考えられる。

このように、事故発生リスクの高い区間の特定や対策の検討を行うためには、該当区間の対策状況などの情報と併せて、活用することが望ましいと考える。

参考文献

- 1) 渡部数樹, 中村英樹: 道路交通環境に着目した交通事故発生要因に関する統計モデル分析. 土木学会論文集, 71 巻, 5 号, pp.I_889-I_901, 2015.
- 2) Mark Poch, Fred Mannering. Negative Binomial Analysis of Intersection-Accident Frequencies, *Journal of Transportation Engineering*, Vol.122, pp.105-113, 1996.
- 3) Kazushi SANO, Touru INAGAKI, Jouji NAKANO, Nguyen Cao Y. An Analysis on Traffic Accidents on Undivided Expressway in Cold and Snow Area, *Eastern Asia Society for Transportation Students*, vol.8, 2010.
- 4) MANH Bui Tuen, PERERA KPD Frank, SANO Kazushi, TKAKURA Takumi, KATO tepei: 高速道路で積雪条件下の交通事故傾向を予測する確率モデルの開発, 土木計画学研究・講演集, vol.66, 2022.
- 5) 四辻裕文, 星卓見, 平澤匡介, 佐藤昌哉: 北海道の高速道路と一般道の平行ルートにおける冬期交通事故リスクとその要因, 寒地土木研究所月報, No.808, 2020.
- 6) 戸澤颯太, 佐野可寸志, 加藤哲平, 原山哲郎, 上石勲: Light GBM を用いた冬期高速道路の路面状況推定, 土木計画学, 2024.
- 7) Lambert, D. Zero-inflated Poisson Regression with an Application to Defects in Manufacturing, *Technometrics*, Vol.34, pp.1-14, 1992.
- 8) 下野嘉子: R を用いた一般化線形モデル (回帰係数編): カウントデータを例に, 雑草研究, 55 巻, 4 号, pp.287-294, 2010.
- 9) search.r-project.org: Dispersion Test, URL: <https://search.r-project.org/CARN/refmans/AER/html/dispersiontest.html>. (参照 2025/01/15)
- 10) search.r-project.org: Vuong test, URL: <https://search.r-project.org/CARN/refmans/unitquandr-eg/html/vuong.test.html>. (参照 2025/01/15)
- 11) 松尾精彦: Vuong test とその正規モデルへの適用法, 関西大学経済論集, 54 巻, 1 号, pp.39-60, 2004.