

報告漏れを考慮した交通事故データの解析*

Analysis on Traffic Accident Data Accounting for Underreporting Bias*

山本俊行**・端地純平***

By Toshiyuki YAMAMOTO**・Junpei HASHIJI***

1. はじめに

交通事故による身体損傷程度について、道路、運転者、車両のそれぞれが及ぼす影響を明らかにするために、これまで交通事故データを用いて数多くの分析が行われてきた。交通事故データは、損傷程度を死亡、重傷、などの離散的な値として記録されていることが一般的である。損傷程度の離散性を考慮した分析手法として、近年ではロジスティック回帰モデルの他、オーダード・プロビットモデルや多項ロジットモデル等の離散選択モデルの適用が進んでいる。

このうち、ロジスティック回帰モデルは損傷程度を二分割することで、説明要因の影響を分析するものであるが、複数のロジスティック回帰を用いた階層的ロジスティック回帰モデルによって、より詳細な損傷程度を用いた分析が可能である。ただし、階層間の独立性の仮定について批判も見られる。また、オーダード・プロビットモデルは全ての損傷程度が同一の尺度で説明できると仮定したモデルであり、階層間の独立性に関する問題はないものの、説明要因が全ての階層（例えば、車両損傷のみか軽傷か、と重傷か死亡か）で同一の影響を持つ仮定に対する批判もある。

一方、離散選択モデルはモデル構築上の柔軟性は高いものの、損傷程度の序列性、すなわち、軽傷の次の損傷程度は重傷、重傷の次の損傷程度は死亡、といった複数の損傷程度の間順番を考慮していないというモデル構造上の本質的な問題を抱えている。本研究では、同一のデータを対象として、階層的ロジスティック回帰モデルとオーダード・プロビット

モデルを構築し、両モデル間の比較を行い有効性を明らかにすることを目的とする。ただし、誤差項の分布形に関する仮定が影響を及ぼすことを避けるため、前者のモデルについては正規分布を仮定した階層的プロビットモデルを用いた分析を行う。

従来の交通事故解析の多くは警察や病院などに報告された事故データをもとに分析されている。しかし、それらのデータの問題点として、全ての交通事故が含まれている訳ではないということが挙げられる。すなわち、特に損傷程度の軽い事故について、警察や病院に報告されないことがある。よって、報告された事故データは全交通事故からの標本抽出と見なすことができる。さらに、その抽出率は全交通事故に対して一定ではなく、損傷程度の軽い事故については抽出率が低いため、選択肢別抽出されたデータと見なすことができる。

一般的には、選択肢別抽出による標本の分析では重み付けによる修正が必要であるが、抽出率が不明な場合の推定方法について実データへの適用事例はほとんど無い。本研究では、Cosslettにより提案された推定方法¹⁾を用い、事故の報告漏れによる推定結果に与える影響、及び、Cosslettによる推定方法の有効性と妥当性について検証することを目的とする。

2. データの概要

本研究で用いるデータは米国ワシントン州で1993年から1996年に発生した交通事故で州道路局に交通事故データとして記録されている20,363件の路側障害物との接触事故である。このデータは警察によって収集されたものであり、最重傷者と運転者の損傷程度が0：車両損傷のみ、1：損傷の可能性あり、2：軽傷、3：重傷、4：死亡、の5段階で記録されている他、路側障害物の種類、天候、時刻等の道路状況、シートベルト、飲酒等の運転状態に関す

*キーワード：交通安全，不完全データ

**正員，博(工)，名古屋大学大学院工学研究科
(愛知県名古屋市千種区不老町，TEL:052-789-4636，
E-mail:yamamoto@civil.nagoya-u.ac.jp)

***正員，修(工)，長大(東京都中央区日本橋蛸殻町
一丁目20番4号，TEL: 03-3639-3301)

表-1 損傷程度の分布

	都市部		郊外部	
	サンプル数	%	サンプル数	%
車両損傷のみ	6125	63	6514	61
損傷の可能性有り	1646	17	1357	13
軽傷	1602	16	2191	21
重傷	297	3	474	4
死亡	53	1	104	1
合計	9723	100	10640	100

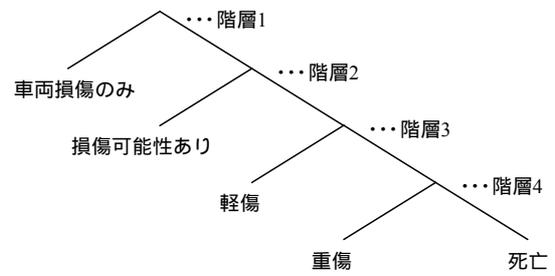


図-1 階層的プロビットモデルの構造

る情報を含んでいる。本研究では、最重傷者の損傷程度を被説明変数としたモデルは異なる乗車人数間で安定性を持たないという著者らの先行研究²⁾の知見に基づき、運転者の損傷程度を対象とした分析を行う。また、同様に先行研究の知見に基づき、都市部と郊外部を区別した分析を行う。運転者の損傷程度の分布を表-1に示す。表より、都市部、郊外部ともに、車両損傷のみの事故が最も多く、重傷及び死亡事故の割合は非常に小さいことが分かる。

3. モデル

(1) オーダード・プロビットモデル

オーダード・プロビットモデルでは、潜在的な連続変数(この場合は実際の損傷程度) y_i^* を仮定し、以下の式で表す。

$$y_i^* = \beta x_i + \varepsilon_i \quad (1)$$

ただし、 β は未知パラメータベクトル、 x_i は説明変数ベクトル、 ε_i は標準正規分布に従う誤差項。

事故データに記録される損傷程度、 y_i は以下で与えられる。

$$\begin{aligned} y_i &= 0 \text{ if } y_i^* \leq 0, \\ &= 1 \text{ if } 0 < y_i^* \leq \mu_1, \\ &= 2 \text{ if } \mu_1 < y_i^* \leq \mu_2, \\ &\vdots \\ &= J \text{ if } \mu_{J-1} < y_i^* \end{aligned} \quad (2)$$

ただし、 μ は未知パラメータ、 J は損傷程度のカテゴリ数。よって、損傷程度 j_i が記録される確率は、以下で表される。

$$\begin{aligned} \Pr(y_i = j_i) &= \int_{\mu_{j_i-1} - \beta x_i}^{\mu_{j_i} - \beta x_i} \phi(\varepsilon_i) d\varepsilon_i \\ &= \Phi(\mu_{j_i} - \beta x_i) - \Phi(\mu_{j_i-1} - \beta x_i) \end{aligned} \quad (3)$$

ただし、 ϕ 、 Φ は標準正規確率密度関数、累積分布関数、 μ_1 、 μ_0 、 μ_J はそれぞれ $-\infty$ 、 0 、 ∞ 。

(2) 階層的プロビットモデル

階層的プロビットモデルはオーダード・プロビットモデルと同様に、連続変数と仮定される実際の損傷程度を序列性のあるカテゴリーによって観測するモデルであるが、損傷程度を決定する関数が階層によって異なることを仮定したモデルである。従来の研究では、損傷程度の低い階層から順に二分割する方法と、損傷程度の高い階層から順に二分割する方法が用いられているが、いずれの方法を用いた場合でも同程度のモデルの適合度が得られている。本研究では、後述の理由により損傷程度の低い階層から二分割する方法を用いる。本研究で用いる階層構造を図-1に示す。また、損傷程度 j_i が記録される確率は、以下で表される。

$$\begin{aligned} \Pr(y_i = 0) &= 1 - \Phi(\beta_{1+} x_i) \\ \Pr(y_i = 1) &= \Phi(\beta_{1+} x_i) \{1 - \Phi(\beta_{2+} x_i)\} \\ \Pr(y_i = 2) &= \Phi(\beta_{1+} x_i) \Phi(\beta_{2+} x_i) \{1 - \Phi(\beta_{3+} x_i)\} \\ \Pr(y_i = 3) &= \Phi(\beta_{1+} x_i) \Phi(\beta_{2+} x_i) \Phi(\beta_{3+} x_i) \{1 - \Phi(\beta_{4+} x_i)\} \\ \Pr(y_i = 4) &= \Phi(\beta_{1+} x_i) \Phi(\beta_{2+} x_i) \Phi(\beta_{3+} x_i) \Phi(\beta_{4+} x_i) \end{aligned} \quad (4)$$

ただし、 β_{1+} 、 β_{2+} 、 β_{3+} 、 β_{4+} は未知パラメータベクトル。通常の階層的プロビットモデルでは階層間の独立性を仮定するため、個々の階層毎にパラメータの推定が可能である。本研究では、事故の報告漏れに着目しており、損傷程度の低い階層から二分割した場合、損傷程度の低い階層のみに報告漏れを含むとすれば、損傷程度の高い階層でのパラメータの推定に影響を及ぼさないという利点を持つ。これが本研究で損傷程度が低い階層から二分割する方法を用いる理由である。

前述のように、階層的プロビットモデルの適用には、階層間の独立性を検証する必要がある。本研究では、以下の式で表される2変量プロビットモデルを構築し、連続する2つの階層間の誤差相関を検証する。

表-2 モデルの適合度

	オーダード・プロビットモデル		階層的プロビットモデル	
	最終尤度	AIC	最終尤度	AIC
報告漏れ考慮無し				
都市部	-9200	9235	-9097	9146
郊外部	-10292	10338	-10191	10261
報告漏れ考慮有り				
都市部	-9193	9230	-9094	9145
郊外部	-10288	10336	-10212	10284

$$\begin{aligned} \Pr(y_i = j-1 | y_i \geq j-1) &= 1 - \Phi(\beta_j x_i) \\ \Pr(y_i = j | y_i \geq j-1) &= \Phi_2(\beta_j x_i, -\beta_{j+1} x_i, -\rho_{j,j+1}) \\ \Pr(y_i \geq j+1 | y_i \geq j-1) &= \Phi_2(\beta_j x_i, \beta_{j+1} x_i, \rho_{j,j+1}) \end{aligned} \quad (5)$$

ただし、 Φ_2 は 2 変量標準正規累積分布関数、 ρ は相関係数。

(3) 報告漏れの考慮

選択肢別抽出標本に対する推定方法としては、WESML 推定量が用いられることが多い。しかしながら、WESML 推定量は各選択肢毎の抽出率が未知、すなわち、母集団内の各選択肢の比率（母比率）が未知の場合には適用できない。Cosslett は母比率が未知の場合の推定量として、以下を提案している。

$$\max_{(\theta, Q)} \frac{1}{N} \sum_{i=1}^N \ln \frac{P(y_i = j_i | \theta) H(j_i) / Q(j_i)}{\sum_{k=1}^J P(y_i = k | \theta) H(k) / Q(k)} \quad (6)$$

ただし、 θ は未知パラメータベクトル、 N はサンプル数、 $H(j)$ は選択肢 j のサンプル内比率、 $Q(j)$ は選択肢 j の母比率（未知）。式(6)では、母比率を未知変数として取り扱っており、未知パラメータベクトルを推定すると同時に、母比率も推定する事が可能である。

4. 推定結果

始めに、報告漏れを考慮せず推定した場合の最終尤度及び情報量基準（AIC）を表-2 に示す。報告漏れを考慮しない場合、都市部、郊外部ともに階層的プロビットモデルの方が最終尤度、AIC とともに優れた値を示しており、オーダード・プロビットモデルより損傷程度の説明力が高いと判断できる。

また、2 変量プロビットモデルによる階層間の誤差相関の推定結果を表-3 に示す。表より、都市部、

表-3 階層間の相関係数の推定値

	都市部		郊外部	
	推定値	t 値	推定値	t 値
階層 1 と 2	-0.299	-2.40	-0.450	-4.37
階層 2 と 3	-0.250	-1.00	-0.330	-1.40
階層 3 と 4	-0.568	-1.35	-0.378	-1.82

郊外部ともに階層 1 と 2 の間に負の有意な誤差相関があることが示された。すなわち、階層 1 において説明変数に含まれていない要因によって損傷程度が損傷の可能性あり以上に分類された場合には、階層 2 において損傷の可能性ありに分類される可能性が高いことを意味している。この原因としては、車両損傷のみと損傷の可能性ありの 2 つの損傷程度の区別が曖昧であり、2 つの損傷程度間の分類が事故を検分した警察官の主観によるところが大きいことが影響している可能性がある。他の階層間では誤差相関が有意となっていないため、車両損傷のみと損傷の可能性ありの 2 つの損傷程度を統合し、損傷程度を 4 段階で表現すれば階層的プロビットモデル適用に問題はないものと考えられる。

報告漏れを考慮する場合、どの損傷程度に報告漏れが存在するかを先験的に決定する必要がある。本研究では、損傷程度の低いものから「車両損傷のみ」だけに報告漏れが存在する場合、「車両損傷のみ」と「損傷の可能性あり」に報告漏れが存在する場合、...、「死亡」以外の全ての損傷程度に報告漏れが存在する場合、の 4 パターンについてモデルの推定を行った。このうち、損傷程度が低いほど報告漏れが発生しやすい、という仮説に合致する結果が得られたのは、都市部については、「車両損傷のみ」と「損傷の可能性あり」に報告漏れが存在すると仮定した場合であり、郊外部では仮説に合致する結果は得られなかった。

都市部の推定結果を表-4 に示す。表より、階層的プロビットモデルにおいてはサンプル数の多い階層 1 においてオーダード・プロビットモデルと同様に多くの説明変数が有意となっているものの、サンプル数の少ない階層において有意な変数が少なくなっていることが分かる。ただし、ラップ&ショルダーベルトはいずれの階層でも有意となっており、シートベルトの重要性が確認される結果となった。

表-4 影響要因のパラメータ推定結果

変数名**	オーダード・プロビットモデル		階層的プロビットモデル							
	推定値	t値	階層 1		階層 2		階層 3		階層 4	
定数項	0.485	1.59	1.203	1.25	-1.667	-17.06	-0.745	-11.8	-1.475	-6.20
道路										
交差点	-0.113	-2.80	-0.147	-2.82						
交差点外	-0.347	-2.93	-0.410	-2.37						
凍結積雪	-0.166	-4.67	-0.182	-3.33	-0.107	-2.29				
道路外	0.518	5.94	0.677	5.64						
雨	-0.117	-4.50	-0.123	-2.78	-0.108	-3.03				
路側障害物										
標識	-0.375	-5.68	-0.463	-4.44						
排水溝	-0.244	-4.51	-0.211	-2.56	-0.309	-4.05				
ガードレール端	0.204	2.23	0.244	1.86					1.257	2.56
ガードレール正面	-0.149	-4.38	-0.169	-3.16	-0.101	-2.29				
コンクリート塀	-0.112	-4.09	-0.109	-2.80	-0.132	-3.82			0.411	2.07
橋	-0.110	-2.21	-0.135	-2.01						
木	0.113	1.83							1.023	3.46
フェンス	-0.240	-3.11	-0.326	-3.02						
車両										
ラップベルト	-0.647	-8.46	-0.778	-3.77			-0.666	-3.83		
ショルダーベルト	-0.489	-3.79	-0.689	-3.20						
ラップ&ショルダーベルト	-0.836	-12.3	-0.963	-4.31	-0.374	-9.82	-0.373	-5.23	-0.578	-3.14
エアバッグ	-0.372	-4.58	-0.394	-2.90						
車齢	-0.006	-3.29	-0.007	-3.02						
トラック大	-0.603	-2.17	-0.73	-2.14						
トラック小	-0.159	-1.74	-0.275	-2.51						
バイク	-0.502	-2.09	-0.683	-1.99						
欠陥ブレーキ	-0.225	-2.10	-0.272	-1.99						
パンク	-0.205	-2.11	-0.297	-2.50						
運転者										
男性	-0.215	-7.60	-0.334	-5.19	0.123	3.97				
年齢	0.001	1.80							0.012	2.23
制限速度超過	0.161	2.47								
速度超過	0.075	2.78	0.072	2.01						
サンプル数		9723		9723		3598		1952		350
初期尤度		-15648.6		-6739.4		-2493.9		-1353.0		-242.6
最終尤度		-9193.3		-5799.6		-2269.2		-892.6		-132.7

5. おわりに

本研究では、Cosslettの方法を用いて報告漏れを考慮した交通事故の損傷程度のモデルを構築した。推定された報告漏れ率等についての推定結果は学会発表時に報告したい。

参考文献

- 1) Cosslett, S.: Efficient estimation of discrete-choice methods, In C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Choice Data with Econometric Applications*, Cambridge, MA: MIT Press, pp. 51-111, 1981.
- 2) Yamamoto, T. and Shankar, V.N.: Bivariate ordered-response probit model of driver's and passenger's injury severities in collision with fixed objects, *Accident Analysis and Prevention*, Vol. 36, pp. 869-876, 2004.