IV—15

対数線形モデルによる冬期交通事故の分析

北海道大学工学部　正員　藤原　隆
同上　正員　加来　照俊

1．まえがき

交通事故の増加にともない、第２次交通戦争が宣言され、効果ある対策の緊急な実施が望まれているが、困難な点も多いのが現実である。また、北海道などの積雪寒冷地では、スパイクタイヤの問題に関連して、新たな冬期道路交通の環境づくりに取り組まなければならない状況にある。

本論では、北海道警察の協力によって行われた冬期の札幌市における交通事故調査データをもとにして、カテゴリカルなデータの要因分析に用いられている対数線形モデルによって、冬期の交通事故の分析を試みようとするものである。事故のデータは、一般に二つ以上の変数間のクロス表の形で表されることが多いが、さらに一歩踏み込んで、変数間の関連性について考察しようとする時、従来の方法では特に多くの変数、要因が関係する事故データに対しては、分析がなかなか困難であると考えられる。その点に関して対数線形モデルは、次節で述べるように都合の良い点があり、その限界や適用に注意しながら用いれば有効な分析が可能ではないかという観点から調査データを用いて例示しようとするものである。

2．モデルについて

交通事故データとして、得られる多くの物は、分類で与えられる、カテゴリカルなデータである。今まで、種々の方法を用いて、交通事故の分析が行われてきたが、本論で用いようとするのは、社会学、医学、工学等の多方面で最近使われている、分割表の分析法の一つである対数線形モデルによるものである。分割表の形で与えられる変数間の関連性については、χ^2 値による独立性の検定などが行われてきたが、対数線形モデルでは、このようなデータの多変量における２個以上の変数間の連関を、系統的に、統一的に構造化された手続きで扱うことができる[3]、という特徴がある。また、分散分析に類似しており離散的なデータに関して、要因の主効果や、要因間の交互作用を検討することができる。

対数線形モデルに関する専門書も多数ある（参考文献２、３、４、５など）ので、理論などに関する、詳細な点についてはそれらにゆずることにする。　このモデルは、繰り返し計算の必要から、計算機の発達とともにますます使われるようになってきたようであり、計算プログラムも参考文献１に示すBASIC で書かれたパーソナルコンピュータで扱えるものや、SAS などの、プログラムパッケージの中のプロシジャーを用いて大型計算機で扱えるものといろいろなものがあるが、ここでは誰でも使うことのできる参考文献１に示されたBASIC のプログラムを用い、１．で述べた交通事故データを対象に考えてみようと思う。

表１　変数Ａ、ＢのＩ×Ｊ表

A＼B	B_1	・・B_j	・・B_J	計
A_1	n_{11}	・・n_{1j}	・・n_{1J}	$n_{1.}$
・	・	・・・	・・・	・
A_i	n_{i1}	・・n_{ij}	・・n_{iJ}	$n_{i.}$
・	・	・・・	・・・	・
A_I	n_{I1}	・・n_{Ij}	・・n_{IJ}	$n_{I.}$
計	$n_{.1}$	・・$n_{.j}$	・・$n_{.J}$	$n_{..}$

Winter Traffic Accident Analisis by Loglinear Model
by Takashi Fujiwara and Terutoshi Kaku

今、二つの変数、A、B（カテゴリー数I、J）を考えると、その分割表は、表1のように表せる。以下、末尾にあげた参考文献をもとにして、モデルについて、簡単に述べる。

変数A、Bのそれぞれi、j番目のセルの観測度数n_{ij}の期待度数を、$m_{ij}=E(n_{ij})$とすると、分割表のデータは、サンプリングの仕方によって、ポアソン分布や多項分布など想定される確率分布が異なるが、今考えている一定期間における交通事故件数などの場合は、ポアソン分布に従うと考えられる。従って、確率は$f(n_{ij})=\prod_i\prod_j(m_{ij}^{n_{ij}}e^{-m_{ij}}/n_{ij}!)$ となる。ここで、二つの変数が互いに独立ならば、次式$m_{ij}=m_{i.}m_{.j}/n_{..}$ の関係が成り立つ。この両辺の自然対数をとると、次式のようになる。

$$\ln(m_{ij})=\ln(m_{i.})+\ln(m_{.j})-\ln(n_{..})$$

そこで、次のように置き換えて、パラメータを導入して；

$$u=\sum_i\sum_j\ln(m_{ij})/IJ、$$
$$u_i^A=\sum_j\ln(m_{ij})/J-u、$$
$$u_j^B=\sum_i\ln(m_{ij})/I-u$$

上式を書き直すと、

$$\ln(m_{ij})=u+u_i^A+u_j^B$$

ここでu_i^A、u_j^Bは、$\sum_i u_i^A=\sum_j u_j^B=0$を満たす。

ここで、分散分析との類似から、u、u_i^A、u_j^Bをそれぞれ、全平均効果、変数Aのi番目のカテゴリーによる主効果、変数Bのj番目のカテゴリーによる主効果とよぶ。また、この式に、変数AとBが独立でない場合の交互作用効果u_{ij}^{AB}を加えたものを飽和モデルとよぶ（カテゴリー数I，Jに関する和が0になる制約条件あり）。　ここで考える三つの変数A、B、C（カテゴリーI、J、K）に対しても、同様の式

$$\ln(m_{ijk})=u+u_i^A+u_j^B+u_k^C+u_{ij}^{AB}+u_{ik}^{AC}+u_{jk}^{BC}+u_{ijk}^{ABC}$$

ができる（ただし、二変数の時と同様なカテゴリーに関するパラメータの和が0になる制約条件がある）。上の式を飽和モデルと呼ぶ。さらにここでは階層モデル（上式で、より高次の項が0でなければ、それより低次の項も式中に含まれていなければならないモデル）のみを考える。

以上のことから、全平均、主効果、交互作用のいづれを含むかによって、種々のモデルができ、それによって変数（要因）間の連関について、考察することができる。

3．対象データについて

例示対象とする冬期交通事故調査データは、札幌市内で1985年から1989年までの 4年間の、それぞれ12月から 3月までに起きた人身および物損事故について得られたが、ここでは最も新しい1988年12月から1989年 3月までのデータを対象に考えることとした。調査された内容の概要は、表2に示すとおりである。

今回は、このうち道路形状、冬型事故、タイヤ種別の三つの要因をとりあげ、上記のデータから、表3のような分割表をつくり、2で述べた方法、参考文献1のプログラムによって、モデルをつくってみることにする。
なお、タイヤに関しては、とりあげた二つのものが他に比べて、非常に多数だったため、2つだけをとりだしてみた。

表　2　　調査内容

分　類	項	目	
自動車関係	駆動方式	タイヤ	車　種
道路　関係	道路形状	道路線形	路面状態
人　　関係	行動目的	行動類型	違　反
事故　関係	事故直前速度	事故類型	冬型事故
そ　の　他	発生年	月日時	天　候

4．モデルの適用結果

今回、とりあげた変数（要因）について、そのカテゴリーの数および内容は、次のようである：

Ａ：タイヤ　（２つ：　スパイク、スタッドレス）

Ｂ：道路形状（４つ：　交差点、交差点付近、単路直線、単路カーブ）

Ｃ：冬型事故（４つ：　スリップ、視界不良、わだち、その他）

これらは、いずれも名義尺度である。

表　3．　三次元分割表

タイヤ	道路形状＼冬型事故	スリップ 観測	スリップ 期待	視界不良 観測	視界不良 期待	わだち 観測	わだち 期待	その他 観測	その他 期待
スパイク	交差点	952	949.08	214	216.01	25	27.77	3887	3885.14
	交差点付近	878	897.24	38	40.11	60	57.02	1879	1860.63
	単路カーブ	363	358.29	16	14.90	49	49.77	298	303.03
	単路直線	1017	1005.38	99	95.99	348	347.44	2496	2511.19
スタッドレス	交差点	238	240.92	39	37.00	7	4.23	622	623.86
	交差点付近	281	261.76	10	7.89	7	9.98	325	343.37
	単路カーブ	70	74.71	1	2.10	7	6.23	45	39.97
	単路直線	237	248.62	13	16.01	51	51.56	408	392.81

（注）表内で、「観測」とは観測値、「期待」とは下に示したモデルによる期待頻度である。

表　4　　モデルによる変数Ｃの主効果とＡ、Ｂとの交互作用のパラメータ推定値

効果	カテゴリ	k＝1	2	3	4
u_k^C		1.18684	-1.54318	-1.34007	1.69641
u_{jk}^{BC}	j=1	-0.18647	0.89912	-1.19951	0.48686
	2	0.28958	-0.18856	-0.33392	0.23290
	3	0.28966	-0.45160	0.69931	-0.53736
	4	-0.39276	-0.25896	0.83412	-0.18240
u_{ik}^{AC}	i=1	-0.15375	0.06735	0.03284	0.05356
	2	0.15375	-0.06734	-0.03284	-0.05356

このデータに対して、さきに述べた方法、プログラムを使って解析すると、次の対数線形モデルが求められる：

$$\ln(m_{ijk}) = u + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC}$$

すなわち、全平均効果、３つの要因の主効果、３つの要因のそれぞれ２要因の交互作用の和からなる。モデルの適合度は、赤池の情報量基準ＡＩＣと、カイ自乗値によって求めている。ＡＩＣは、いくつかのモデルの選択基準であり、次式で表される：

ＡＩＣ＝－２×（最大対数尤度）＋２×（推定すべきモデルの自由なパラメータ数）

この場合、いくつかのモデルのうち、ＡＩＣが最小となるものを選んでいる。

このプログラムでは、モデルは、階層性を仮定しているので、上のように全平均効果、主効果、交互作用すべてを含むものから、主効果のみのもの、全平均効果のみのものまでのいくつかがあり、その中からＡＩＣのような基準でもっとも適合性のよいモデルを選択している。

出力として、他に対数線形モデルの各パラメータの推定値も得られるが、表４に要因Ｃに関する出力を示す。

ここで行った例は、冬型事故（スリップ、視界不良、わだち、その他）という類型を説明する要因として、道路形状とタイヤを選んで、うまくあてはまる対数線形モデルを選択したことになろう。出力から、冬型事故と道路形状、タイヤ間におけるカテゴリーによる差の存在を示唆するものがあると考えられる。参考文献５では、期待頻度の見込み（カテゴリーによる期待頻度の比）からロジットモデルによってロジット分析を行い、分析をすすめる例を示している。

５．あとがき

今回は、質的なデータに対する変数間の関連性を検定、推定する方法である対数線形モデルについて、それを、やはりカテゴリーに分類されて表されることの多い交通事故データに応用し、より系統的、統一的に変数間の関連性について示唆を与えてくれることをめざしたが、実際の適用にあたっては、たとえば参考文献６のなかで例示されているように（一つのミスコードが結果に与える影響）、データや変数、カテゴリーの選択、交絡因子などの点において、慎重な検討が必要である。今回は、方法の例示にとどまり、内容についてのより慎重な分析は今後に残すこととなった。このような方法は、決定的な結果を得るというよりも、モデルで示唆されたことを足がかりとして、分析をすすめていく、探索的な方法ともいえると思う。その意味からも、長所を生かした適用によって、有効に活用したいと思う。

６．参考文献

１．田中、垂水、脇本：パソコン統計解析ハンドブックⅡ．多変量解析編、共立出版、1984
２．柳川　堯：離散多変量データの解析、応用統計学シリーズ、共立出版,1986
３．アプトン：調査分類データの解析法、朝倉書店、1980
４．エヴェリット：質的データの解析ーカイ自乗検定とその展開、新曜社、1980
５．松田紀之：質的情報の多変量解析、朝倉書店、1988
６．Harry S.Lum :Loglinear Models in Traffic Studies, PUBLIC ROADS Vol.52,No.4 1989
７．緒方、松原、柴田：多重環境要因の複合効果の統計解析ー対数線形モデルによる分析、応用統計学、Vol 13,No 3,1984
８．SAS Users' Guide, SAS Institute Inc.
９．岸本　淳司：北海道ＳＡＳユーザー会資料、1989