

GLMは日本に定着するか

オーガナイザー	岩沢 宏和
パネリスト	
チューリッヒ	梅田 和彦
全共連	新居 悠輝
タワーズワトソン	高橋 伸

【岩沢】 時間になりましたので、パネルディスカッション「GLMは日本に定着するか」という題で、お話をしていきます。司会の岩沢です。どうぞよろしくお願いいたします。

まず、いきなりなのですが、われわれは失敗したかと思っている点があります。この時間、パネルディスカッションを90分で申請したのですが、準備してみますと、もっと長く語りたいたいところがありました。そこで少しでも時間捻出のため、司会者がいろいろ話しても仕方がないので、私は、すみません、お聞きづらいたと思いますが、早口で進めていきます。途中、双方向ツールを使いまして会場のみなさんにお聞きするところがありますが、そこも早めに反応していただくと運営上大変ありがたいので、ご協力をお願いいたします。

要旨

- 海外で損害保険の料率算出等で広く用いられている**GLM**について、日本での活用事例を紹介するとともに、今後の発展の可能性についてディスカッションする。
- 日本における**GLM**の実践者である3人の提題者から発表のあと、実際のところ日本で**GLM**をうまく活用するためには、理論的な面や技術的な面でどういったことが重要なのかについて、掘り下げたディスカッションを行う。

3

テーマは、GLM、一般化線形モデルということですが、これはすでに海外では主に損害保険の料率算出等で広く用いられていますが、日本でも活用事例がありますので、その紹介をするとともに、そうはいっても日本ではなかなかまだ用いられていないようだというので、今後の発展の可能性についてディスカッションをしていきたいと思っております。

そこで今日は、日本においてGLMの実践をすでに行われている方々である3人の提題者から発表していただきます。そのあと、実際のところ日本でGLMをうまく活用するためには、どのような論点があるのかについて、ディスカッションを進めていきます。

提題者(発表順)

- 梅田和彦(チューリッヒ保険)
日本におけるこれまでのGLM導入の歴史を概観し、同時にGLMの基本的な特徴を確認する。
- 新居悠輝(全共連)
より具体的な事例におけるGLM導入の背景を説明し、実際のモデリングの例と成果とを紹介する。
- 高橋伸(タワーズワトソン)
海外での事例を含むその他のGLM活用の可能性、モデリングの際の一般的注意点などを紹介する。

私は結構臨機応変が好きでして、それは聞こえはいいかもしれませんが提題者にとってはつらいところ
です。というのも、スライドでは「発表順」となっていますが、これは、実は双方向ツールの結果によって
決めるつもりでして、ここにあるのは仮のものです。ただ、順番はともかく梅田さん、新居さん、高橋さん
に発表していただくことはたしかです。

現在チューリッヒ保険の梅田さんはいくつもの外資系の会社ですでに実際にGLMを導入し活用してきた、
日本では先駆者的な方ということでお話をいただきます。GLMの導入の歴史や、あるいは基本的な特徴に
ついてお話をいただきます。

全共連の新居さんは、外資系ではなくて国産のGLMを実務で活用されているということで、実際に苦労
された点も含めて、GLM導入の背景等や、それからモデリングの例と成果等についてご紹介いただきます。

タワーズワトソンの高橋さんは、GLMもちろんですが、保険会社に役立つ最先端のツールに関しては
日本では第一人者と思われ、詳しくご存じですので、海外の事例を含む、GLMのいろいろな活用方法、あ
るいはモデリングの際の一般的な注意点などについてお話をいただきます。

先ほどお話ししましたが、双方向ツールを使用いたします。その際、個々の質問に回答いただき、
——GLMをやる人がこうして得られる情報を互いにリンクさせないで使うのはどうかと思うのですけれど
も——今回は単にそれぞれの回答割合を把握するためだけに使います。ただ、その後、ジャーナルか会報か
分からないですけれども(編注:実際は「年次大会報告集」)、そこにもその結果とともに報告するという形
で使用いたしますので、ご理解のうえ、ぜひご協力いただければと思います。

双方向ツールの使用目的、範囲

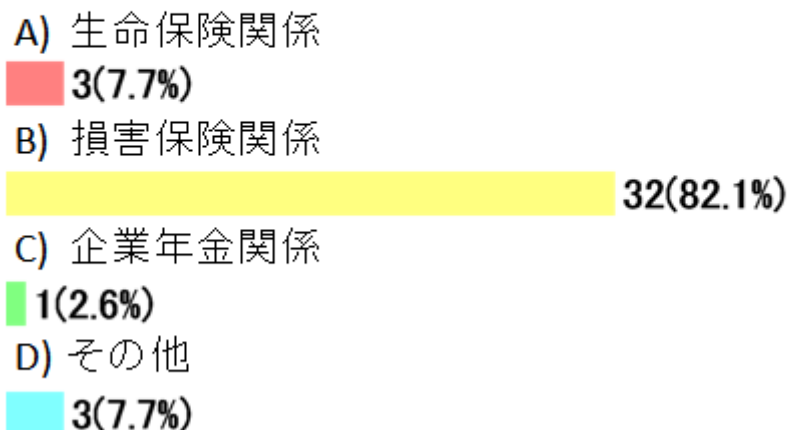
- 本パネルディスカッションでは、双方向ツール（選択形式での質問に本ツール端末所持者からご回答いただく）を利用します。
- 本ツールにより得られた情報は、各選択肢の回答人数および割合（分母は実回答者数とする）を把握するためだけに使用し、その結果は、後にジャーナルまたは会報にも掲載されます。
- 質問は、パネル発表の前にも後にも行います。

3

質問1は、予備的な質問なのですが、お答えいただきたいと思います。実務でのGLMの使用状況を把握したいと思いますので、その予備情報として、どのような業態の方々がこの会場にいらしているのか、会社名というよりも、何の仕事をしているかということでお答えいただければと思いますが、A、B、C、D、生保なのか、損保なのか、企業年金関係なのか、あるいはその他なのかということで、今からお答えください。はい、もう押してください。素晴らしい。早いですね。ありがとうございます。端末を持った方が全員回答されたかどうか分からないのですが、損保関係がかなり多いということですね。ありがとうございます。

質問1

実務でのGLMの使用状況を把握したいと考えています。その予備情報として、現在携わっている業務をご回答ください。



5

次の質問2にまいります。実務にGLMを使っているかどうかということをお答えいただければと思います。通常業務で使っているのか、特別なときや試験的な目的のみで使っているのか、今は使っていないが使ったことがあるのか、これは他の会社にいたときなども含めてですね。あるいは、業務で使ったことがないか。これをいずれか、答えていただければと思います。今からです。はい、お願いします。はい、ありがとうございます。使ったことがないという方が多いですね。6、9、3、21と。はい、ありがとうございます。

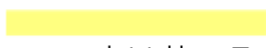
質問2

実務でGLMを使っていますか？


A) 通常業務で使っている

 6(15.4%)

B) 特別なときや試験的な目的のみで使っている

 9(23.1%)

C) いまは使っていないが、使ったことがある(いまとは別業務や別会社の場合を含む)

 3(7.7%)

D) 業務で使ったことはない

 21(53.8%)

7

質問3

前質問でA、BまたはC(GLMを使っている(いた))と回答した方のみご回答ください。その業務は下記のうちのどれですか？

A) 生命保険関係

0(0%)

B) 損害保険関係

 17(85%)

C) 企業年金関係

0(0%)

D) その他(回答不可の場合を含む)

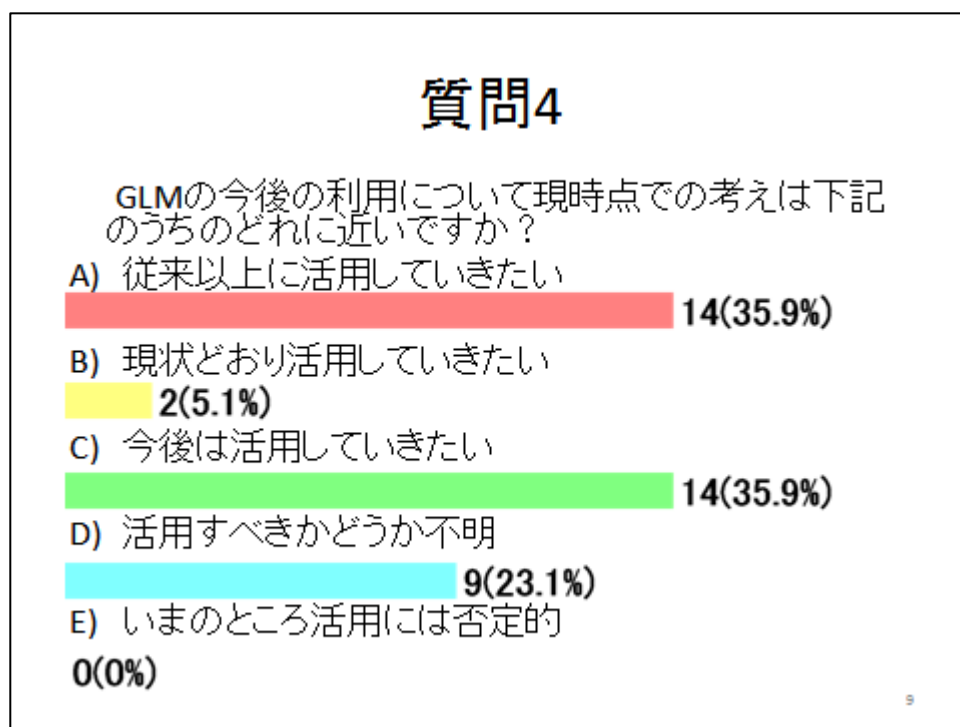
 3(15%)

8

質問3です。使っている、あるいは使っていたと回答された方、ですから、18人いらっしゃると思います

けれども、その仕事はどうだったかということをお答えいただければと思います。お願いします。なぜか先ほどの人数より多いですね、原因は不明ですけども。この回答でも損保関係が多いということでした。

質問4です。今の時点で、どのようなモチベーションでこの会場にいらっしゃるかということ、ぜひ教えてください。これは重要な質問ですので、全員にお答えいただきたいのですが、使用されていたという方は先ほど20名ということですが、従来以上に活用していきたいのか、今のままでいいいいのか、あるいはその他の方は、今後は活用していきたいのか、あるいは活用すべきかどうか不明なのか、今のところは否定的なのだけれどもここに来てみたという感じなのか、お答えいただければと思います。ありがとうございます。



この結果によると、この会場には結構モチベーションの高い方が来ているということで、そのシナリオの場合には、当初の予定どおりの順番で発表をしていきます。提題者は安心しています。

それでは、第1の提題者、梅田さんからお願いいたします。

【梅田】 チューリッヒ保険の梅田でございます。始めにGLMの歴史からご説明いたします。

ここでは三点挙げさせていただきました。第一に、外資系損保によるリスク細分化自動車保険の販売開始が挙げられます。約20年前の話です。当時、日米保険協議の中で、日本の損害保険業界において自動車保険の規制が非常に強いという米国側の要求がありました。この協議の結果、一定の条件の下で、自動車保険につきリスクの細分化が可能になることが決まりました。

翌1997年9月から販売開始ということで、米系の会社が自動車保険のリスク細分化を行ったということがまず背景にあるかと思います。こちらにつきまして、われわれアクチュアリーとしてどのようなことが注目すべき点かと申しますと、全く新しいリスク区分が導入されたことです。当時の自動車保険には年齢区分はありましたが、1) 全年齢担保、2) 21歳未満不担保、3) 26歳未満不担保の3区分しかありませんでした。日米保険協議の結果、年齢区分の細分化、安全装置、例えばエアバッグ、ABSの導入、さらに地域別、走

行距離別、免許証の色別等が新たなリスクファクターとして導入されたのです。

リスクの細分化と、いかに適正な料率水準を保つことができるかというところが、きっかけだったと思います。

日本におけるGLMの歴史

- ▶ 外資系によるリスク細分化自動車保険の販売開始（日米保険協議と1997年9月）
- ▶ 海外におけるOne-wayからMulti-wayへの流れ
- ▶ 料率競争とポートフォリオ安定化の必要性

2

一方、海外に目を転じますと、1990年代後半から、イギリスで通信販売による自動車保険会社が脚光を浴び始めました。また同じ時期に、アクチュアリーファームによるGLMの専用のソフトウェアが開発されました。次第に日本にも話が伝わり始めてきたのです。さらに、そのソフトを導入した会社は、海外では損害率が良好であるということがいわれていました。このようにして、One-way から多変数を一度に解析するというMulti-way への流れがあったと思います。

GLMの例

Generalized Linear Models for Insurance Data (P. Jong and G. Heller)
のデータセット"car"を使って、RによりGLMのモデルを実行した。

変数名	内容
veh_value	車両価格（単位:\$10,000）
exposure	エクスポージャ：0.0027~0.9993（=1/365.25~365/365.25）
dm	クレームの有無：0=無, 1=有
numclaims	クレーム件数
claimcost	クレーム額（クレームがない場合0）
veh_body	車種： BUS, CONV = convertible, COUPE, HBACK = hatchback, HDTOP = hardtop, MCARA = motorized caravan, MIBUS = minibus, PANVN = panel van, RDSTR = roadster, SEDAN, STNWG = station wagon, TRUCK, UTE = utility
veh_age	車齢：1（最若齢）, 2, 3, 4
gender	運転者の性別：M, F
area	運転者住所：A, B, C, D, E, F
agecat	運転者年齢：1（最若齢）, 2, 3, 4, 5, 6

使用したデータはMacquarie Universityのサイトから直接入手できる。

http://www.businessandconomics.mq.edu.au/our_departments/Applied_Finance_and_Actuarial_Studies/research/books/GLMsforInsuranceData

3

料率が競争にさらされると、ポートフォリオの安定化が重要になってきます。リスクを細分化すると、あるものについてはダブルカウント、またトリプルカウントになってくるおそれがあるため、過度の割引または割増を回避することが課題になります。アクチュアリーとして、いかにポートフォリオを安定化させることができるかということで、GLMを使って料率を安定化させることが重要になってきました。

では、GLMのモデリングで一つの例を実行しました。データは、オーストラリアのマッコーリ大学のサイトから入手しました。本のタイトルは『Generalized Linear Models for Insurance Data』です。こちらにあるデータセットはCar(カー)という自動車保険のデータでこちらを使いました。

変数としては幾つかあります。入っているフィールドにつきましては、ビークルバリュー(車両価格)ということで、1万ドル単位で入っています。エクスポージャーの計算は1年間を365.25日としています。うるう年がありますので、1年間をこの日数で考えているようです。

クレームについては、クレームの無いものは0、有ったものは1にしています。また、1事故で複数件クレームが有ったかという情報も入っています。車種については13区分です。次のビークルエイジは車齢で1から4までの区分です。さらに性別で2区分、ドライバーの住所が6区分、ドライバーの年齢区分が6区分あります。

切片項だけのGLMの結果(1)

Call:

```
glm(formula = clm ~ 1, family = poisson, data = car)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.3692	-0.3692	-0.3692	-0.3692	1.8731

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.68613	0.01471	-182.7	<2e-16 ***

...

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 24841 on 67855 degrees of freedom

Residual deviance: 24841 on 67855 degrees of freedom

AIC: 34,091

4

こちらは切片項だけ、すなわち平均だけを取った場合です。つまり、 $y = a + b x$ のときに、 a だけを用いた場合です。なお、分布はポアソンを仮定しています。これを走らせると、切片が-2.68613になっています。赤池情報量(AIC)は34,091です。この例が先ず第一番目のケースですが、今の数値は契約台数をベースにしたモデルでしたので、次に作ったモデルは経過台数をベースにしています。経過台数とは、例えば1月1日から12月31日までの経過期間の場合は1台・年ですが、7月1日から12月31日までの経過期間の場合は0.5台・年になります。

経過台数分の事故件数の期待値の対数を取り、線形予測値を求めます。これを変形すると、オフセット項として調整が必要になります。

契約台数から経過台数の調整

- ログリンクを用いた事故率モデルでは、
- $\text{Log}(E(\text{事故件数})/\text{経過台数}) = \text{線形予測値}$
- すなわち、 $\text{Log}(E(\text{clm})) = \text{線形予測値} + \text{log(exposure)}$ となり、オフセット項による調整が必要。

5

切片項だけのGLMの結果(2)

```
Call:
glm(formula = clm ~ 1, family = poisson, data = car, offset = log(exposure))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5391 -0.4426 -0.3398 -0.2181  3.6957

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.92823    0.01471  -131.1  <2e-16 ***
...
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for poisson family taken to be 1)
Null deviance: 23803  on 67855  degrees of freedom
Residual deviance: 23803  on 67855  degrees of freedom
AIC: 33,053
```

6

ここで変わった点は、最後にオフセットの式が入りました。この結果、切片は-1.92823になり、この結果は後でお見せしますが同じになります。ここで注目すべき点はA I Cが 34,091 から 33,053 に減少したことです。

こちらはケース1の場合が、切片は-2.68613です。ケース2では-1.92823です。この数字は実際に何%かと計算しますと、下段でeの肩に乗せるとケース1は0.068144です。ケース2では0.145405になります。ケース1では、NOVW (Number of Vehicle Written) は67,856台になりますので、これにこれを掛けて

あげて、Number of claim は 4,623.99 です。ケース 2 の方は、経過台数ですので、31,801 台に事故率を掛けて 4,624.01 が得られます。ほぼ同じになりました。ただ、ポイントは、先ほど申しましたけれども、A I C が 34,091 から 33,053 になったということです。

GLM結果の比較

Case1	Intercept	Std Error	Mean-2SE	Mean+2SE	
	-2.68613	0.01471	-2.71555	-2.65671	
	0.068144	1.014819	0.06617	0.07018	
	NOVW	NOC	Mean-2SE	Mean+2SE	
	67,856	4,623.99	4,489.9	4,762.0	AIC: 34,091
Case2	Intercept	Std Error	Mean-2SE	Mean+2SE	
	-1.92823	0.01471	-1.95765	-1.89881	
	0.145405	1.014819	0.14119	0.14975	
	Avg Exp				
	0.468652				
	NOVE	NOC	Mean-2SE	Mean+2SE	
	31,801	4,624.01	4,490.0	4,762.1	AIC: 33,053

7

変数項を含めたGLMの結果(3)

Call:

```
glm(formula = clm ~ veh_body + veh_age + gender + area + agecat,
     family = poisson, data = car2, offset = log(exposure))
```

Coefficients:

```

Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.939074  0.048983 -39.587 < 2e-16 ***
veh_bodyBUS  0.905047  0.335139  2.701 0.006923 **
veh_bodyCONVT -0.515717  0.578124 -0.892 0.372364
veh_bodyCOUPE  0.411678  0.124711  3.301 0.000963 ***
veh_bodyHBACK -0.033629  0.038773 -0.867 0.385780
veh_bodyHDTOP  0.148326  0.082505  1.803 0.108837
veh_bodyMCARA  0.809494  0.288925  2.266 0.023426 *
veh_bodyMIBUS -0.009104  0.155625 -0.058 0.953353
veh_bodyPANVN  0.055058  0.130568  0.422 0.673257
veh_bodyRDSTR  0.094372  0.708047  0.133 0.893968
veh_bodySTMWG  0.064478  0.040460  1.594 0.111019
veh_bodyTRUCK -0.006944  0.097093 -0.072 0.942984
veh_bodyUTE    -0.153981  0.069363 -2.220 0.026425 *
veh_age1      0.082827  0.044399  1.866 0.062108 .
veh_age2     0.111700  0.039276  2.844 0.004455 **
veh_age4     -0.087055  0.040070 -2.173 0.029814 *
genderM      -0.017302  0.031058 -0.557 0.577472
areaA        -0.033021  0.040408 -0.817 0.413819
areaB        0.047367  0.041797  1.133 0.257105
areaD       -0.114383  0.052716 -2.170 0.030023 *
areaE       -0.047866  0.053188 -0.823 0.410727
areaF        0.031389  0.067411  0.468 0.641476

agecat1      0.269983  0.054379  4.963 6.94e-07 ***
agecat2     0.083338  0.044720  1.864 0.062383 .
agecat3     0.031681  0.042646  0.743 0.457548
agecat5    -0.199460  0.050413 -3.956 7.61e-05 ***
agecat6    -0.192285  0.060808 -3.162 0.001566 **

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 23903 on 67855 degrees of freedom
Residual deviance: 23648 on 67829 degrees of freedom
AIC: 32950

Number of Fisher Scoring iterations: 6

```

car2は、carに対しい以下の加工を行ったもの
 ・veh_age, agecatを因子型に変更
 ・veh_body, veh_age, gender, area, agecatについて、エクスポージャー最大の区分の較差が1となるよう因子水準の順序を変更

先ほどは切片項だけでしたが、説明変数を追加しました。今回追加した変数が、車体の種類、車齢、性別、地域と年齢区分の5つです。結果を見るとp値が低いものほど星の数が多く、結果的にはA I Cが 32,950 になりました。

変数項を含めたGLMの結果(4)

- (3)の結果をstepAIC(library MASS)に投入

```
result.step <- stepAIC(result.3)

Call:
glm(formula = clm ~ veh_body + veh_age + area + agecat,
    family = poisson, data = car2, offset = log(exposure))

Coefficients:
              areaD      -0.113268    0.052679   -2.150  0.03154 *
              areaE      -0.047132    0.058173   -0.810  0.41783
              areaF       0.032130    0.067397    0.477  0.63355
              agecat1     0.269990    0.054379   4.965  6.37e-07 ***
              agecat2     0.083973    0.044705   1.878  0.06033 .
              agecat3     0.032242    0.042833    0.756  0.44949
              agecat5     -0.200087    0.050401   -3.970  7.19e-05 ***
              agecat6     -0.194213    0.060709   -3.199  0.00138 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 23803  on 67855  degrees of freedom
Residual deviance: 23648  on 67830  degrees of freedom
AIC: 32948

Number of Fisher Scoring iterations: 6
```

次に stepAIC という関数を使ってA I Cを下げていきました。その結果、性別が抜けA I Cは 32,948 になりました。

変数項を含めたGLMの結果(5)

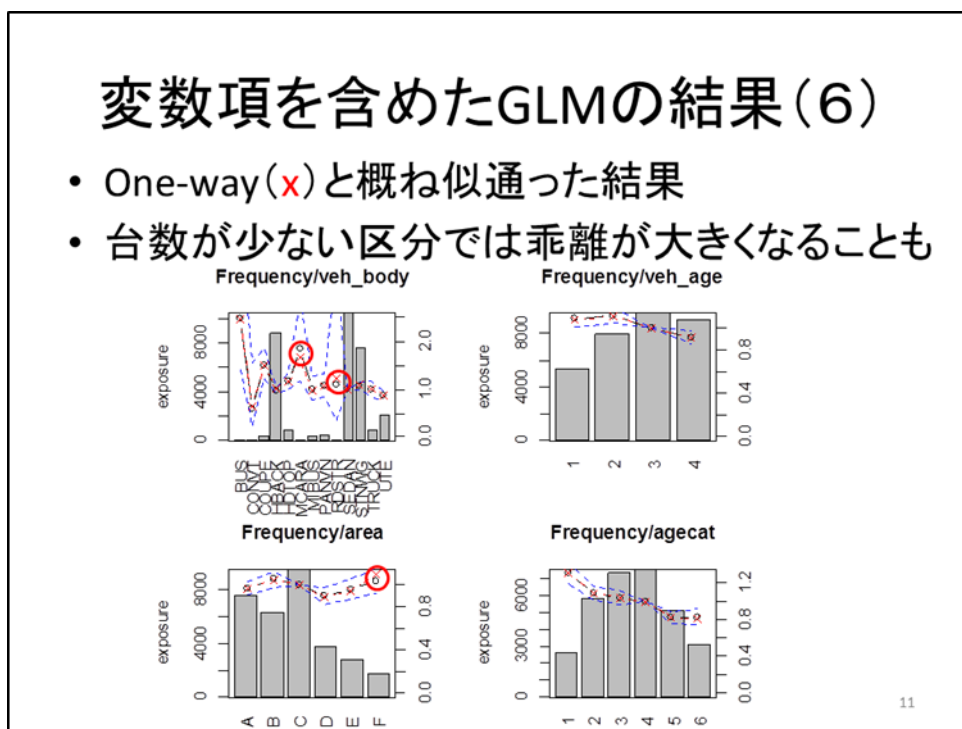
項目	GLM					one-way				
	係数	標準誤差	較差	95%下限	95%上限	エクスポージャー	クラス件数	クラス頻度	較差	
veh_body	BUS	0.901	0.335	2.461	1.276	4.746	26	9	0.348	2.464
	CONVT	-0.515	0.578	0.597	0.192	1.855	33	3	0.092	0.651
	COUPE	0.410	0.125	1.507	1.180	1.924	319	68	0.213	1.508
	HBACK	-0.032	0.039	0.968	0.898	1.045	8,810	1,264	0.143	1.015
	HDTOP	0.145	0.092	1.156	0.965	1.386	783	130	0.166	1.174
	MCARA	0.606	0.269	1.832	1.082	3.103	59	14	0.236	1.671
	MBUS	-0.010	0.156	0.990	0.730	1.343	317	43	0.136	0.960
	PANVN	0.048	0.130	1.049	0.813	1.354	409	62	0.152	1.072
	RDSTR	0.088	0.708	1.092	0.273	4.373	12	2	0.171	1.213
	SEDAN	0.000	0.000	1.000	1.000	1.000	10,445	1,476	0.141	1.000
	STNWG	0.062	0.040	1.064	0.983	1.151	7,638	1,173	0.154	1.087
TRUCK	-0.014	0.096	0.986	0.816	1.190	844	120	0.142	1.006	
UTE	-0.160	0.068	0.852	0.745	0.974	2,106	260	0.123	0.874	
veh_age	1	0.083	0.044	1.086	0.996	1.185	5,339	825	0.155	1.083
	2	0.112	0.039	1.118	1.035	1.208	7,924	1,259	0.159	1.113
	3	0.000	0.000	1.000	1.000	1.000	9,542	1,362	0.143	1.000
	4	-0.088	0.040	0.916	0.847	0.991	8,996	1,178	0.131	0.917
area	A	-0.033	0.040	0.967	0.894	1.047	7,597	1,085	0.143	0.969
	B	0.047	0.042	1.049	0.966	1.138	6,298	965	0.153	1.039
	C	0.000	0.000	1.000	1.000	1.000	9,578	1,412	0.147	1.000
	D	-0.113	0.053	0.893	0.805	0.990	3,820	496	0.130	0.881
	E	-0.047	0.058	0.954	0.851	1.069	2,772	386	0.139	0.945
	F	0.032	0.067	1.033	0.905	1.178	1,736	280	0.161	1.094
agecat	1	0.270	0.054	1.310	1.178	1.457	2,612	496	0.190	1.310
	2	0.084	0.045	1.088	0.996	1.187	5,892	932	0.158	1.091
	3	0.032	0.043	1.033	0.950	1.123	7,409	1,113	0.150	1.036
	4	0.000	0.000	1.000	1.000	1.000	7,617	1,104	0.145	1.000
	5	-0.200	0.050	0.819	0.742	0.904	5,171	614	0.119	0.819
	6	-0.194	0.061	0.823	0.731	0.928	3,100	385	0.118	0.812

ここでは最後の(4)の結果を表にまとめ、GLMと One-way の比較を行いました。上から車種、車齢、地域、年齢区分ですが、それぞれの中でエクスポージャーが最も多いところ、すなわち台数が多い区分をベースにしています。例えば車種につきましてはセダンが一番多いということで、係数は 1.0 にしています。

左側半分がGLMの結果で右側が one-way の結果です。較差につき、one-way の較差とGLMの較差が、どこから違いが出てくるかということなのですが、エクスポージャーが比較的少ないところで現れるようです。例えば、MCARAというのはキャラバンなのですが、台数が59台で少ない区分です。one-wayですと、較差的には1.671倍と結果が出るのですが、GLMでは較差は1.832ということですので、one-wayの方が低く、実際はもっと高いということです。

同じようなことが言える区分が他にも幾つかあります。RDSTR（ロードスター）についても、one-wayの較差1.213に対してGLMの較差1.092ということで、実際は較差が低いということです。

また、Fの地域につきましては、GLMは1.033ですが、one-wayは1.094ということで、そこも乖離が出ています。



今までの結果をグラフ上で示したものです。one-way は x で示しています。概ねGLMと one-way で較差は近いのですが、較差があるところについては、プライシングの観点では気をつけるべきところです。

このようにエクスポージャーの大きさにばらつきがある会社では、リスク細分化にあたりこのようなところを良く見ておく必要があります。私の経験からこのようにして、GLMは役に立つと考えるようになりました。

【岩沢】 ありがとうございます。

それでは、引き続きまして、新居さん、お願いします。

【新居】 全共連の新居と申します。よろしくお願いします。

本日ご紹介させていただく内容は、弊社で数年前に自動車保険のリスク分析のためにGLMを導入したということがありましたので、そのときの経験に基づいたものです。なお、本内容はあくまで私の個人的な見解でして、所属する組織の意見ではありませんことを、あらかじめお断りさせていただければと思います。

それでは、発表に入らせていただきます。

まず、一般化線形モデルの導入の背景についてです。早速なのですけれども、今、皆様が自動車保険のリスク分析を担当しているとし、上司に、ある安全装置の有無がクレームコストに与える影響を分析してほしいと言われたとします。このとき、上司の指示どおりに、その安全装置の有無別に実績クレームコストを集計してみたところ、こちらの表のようになっていたとします。

背景(1)

- 自動車保険で、ある安全装置の有無別のクレームコストが以下の通り得られているとき、この安全装置の影響は？

	安全装置あり	安全装置なし	合計
クレームコスト	17,455	18,077	17,940
契約台数	2,200	7,800	10,000

※ 数字は仮(以下同様)

⇒ $17,455 / 18,077 = 0.965 \dots$ ⇒ 約3%?

3

数字は仮のものですが、クレームコストについて、この安全装置がある場合は17,455円、ない場合は18,077円という結果が得られましたので、この17,455円を18,077円で割ると0.965になるということで、3%ほどクレームコストを低減する効果があるのではないかと思えるわけです。ですので、上司には3%ほどクレームコストを低減する効果が見られますと報告するのですが、そうしますと、今度は上司から、すべての車種でそういう傾向があったのかと聞かれるわけです。そこで、今度は車種別に分けてみることになります。

分けてみたところ、安全装置ありではクレームコストの合計は17,455円で当然変わらないのですが、車種Aでは17,000円、車種Bでは22,000円、車種Cでは安全装置がついているものがない、となっています。また、安全装置なしでは、車種合計では18,077円ですが、車種別に見ると、車種Aでは16,000円、車種Bでは2万円、車種Cでは21,000円となっています。

ですので、車種Aにおいても、車種Bにおいても、安全装置ありの方が、安全装置なしに比べてクレームコストが高い状況になっていることに気づきます。だとすると、先ほど上司に3%ほど効果があると言ってしまったものの、車種別に見ると、実はクレームコストを低減させる効果がないということになりました。

では、この安全装置には本当にクレームコストを低減させる効果がないのでしょうか。結論から言いますと、まだよく分からないということです。といいますのは、安全装置の有無に相関がある項目が、本当に車種だけなのかということを考える必要があるからです。

背景(2)

- さらに、車種別にクレームコストを調べると以下の通りとなっていた。この安全装置の影響は？

	車種	安全装置あり	安全装置なし	合計
クレームコスト	A	17,000	16,000	16,333
	B	22,000	20,000	20,133
	C	-	21,000	21,000
	合計	17,455	18,077	17,940
契約台数	A	2,000	4,000	6,000
	B	200	2,800	3,000
	C	0	1,000	1,000
	合計	2,200	7,800	10,000

⇒クレームコストを低減させる効果はない・・・？

例えば、安全装置がある自動車は製造年が新しいものかもしれませんが、安全装置の有無は自動車の製造年の影響を受けそうですので、車齢と相関があるかもしれません。また、安全装置がついているということは、自動車の価格が高い可能性がありますので、自動車の価格、つまりは車価とも相関があるかもしれません。もしかしたら、そのような安全装置がついている車に乗りたい人たちがいるかもしれませんので、購買層とも関係があるかもしれません。

背景(3)

- 安全装置の有無に相関がある項目は、車種だけ？
 - 自動車の製造年の影響を受けそうなので、車齢とも相関があるかもしれない
 - 自動車の価格(車価)とも相関があるかもしれない
 - もしかしたら、購買層(若者、男性等)とも関係があるかもしれない

⇒多次元のリスク評価には、多次元のモノサシが必要！

このように、いろいろな要素が組み合わさってできている自動車保険のリスク、つまりは多次元のリスク

を評価していくためには、各要素を同時に見ていくためのツール、つまりは多次元の物差しが必要だということを感じた、というのがGLMの導入に至った最も大きな理由でした。ですので、リスクをよく知りたいという、割と純粋な動機からGLMを導入いたしました。

それでは、実際のモデリングの例に入っていきたいと思います。私の経験は、GLMを自動車保険のリスク評価に使ったという事例に限られますので、このケースについてご紹介したいと思います。

2. ケーススタディ

自動車保険のプライシングの補助として

モデリングに有用な文献

- モデリングに先立って文献調査を行い、[1]をバイブルにした

	実務面	理論面
[1]A Practitioner's Guide to Generalized Linear Models (D. Anderson etc)	☆☆☆	☆☆
[2]Non-Life Insurance Pricing with Generalized Linear Models (E. Ohlsson, B. Johansson)	☆☆	☆☆☆
[3]Generalized Linear Models for Insurance Data (P. Jong, G. Heller)	☆☆	☆☆☆
[4]統計モデル入門 (アネットJ.ドブソン)	☆	☆☆☆

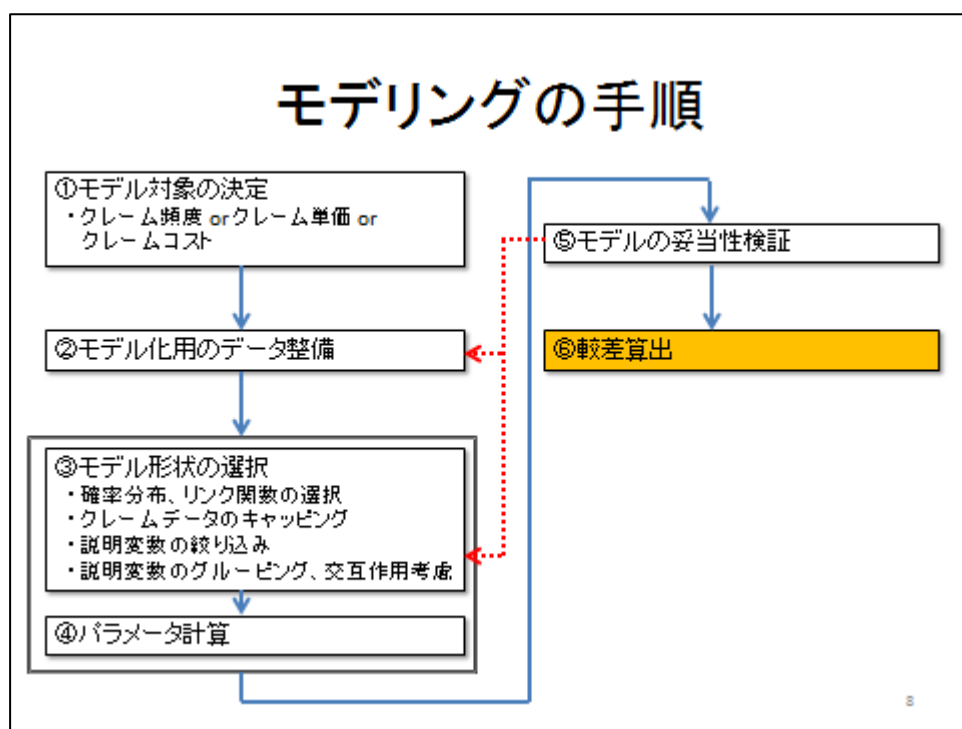
※あくまで主観な評価

7

まず、モデリングに先だって文献調査を行いました。とはいえ、数年前は特にですが、GLMの本自体が少なく、しかも保険データへの適用について書かれている本はほとんどないという状況でしたので、結局あまり選択肢はありませんでした。

当時はまだ、この2番の本は出ていませんでしたので、こちらの表に記載の1番と3番と4番の本を比較しました。表の星印はあくまで主観的なものですが、この1番に挙げている、『A practitioners Guide to GLM』という本が、実務面の記載が非常に豊富でしたので、これをバイブルにしました。

余談なのですが、この1番の本については、アメリカの損保アクチュアリー会のサイトから無料でダウンロードできるようになっています。それから、この2番の本は、今、ASTIN関連研究会が翻訳している本でして、1番と比べますと理論面の記載が豊富で、また、実データによる分析ができるように準備されていますので、その点も非常に有益な本かと思っております。



続いて、モデリングの手順です。個別の手順につきましては後ほどご紹介しますので、ここではモデリングのざっくりした流れをご覧いただければと思います。モデリングは、まずはモデルの対象を決めるところから始まります。続いて、データ整備を行います。そして、モデルの形状を選択して、パラメーターを計算するという流れになります。そうしますと、とりあえずは何らかの結果が算出されますので、それが妥当かどうかを考えまして、何か違うと思ったらモデルの形状の選択をやり直したり、場合によってはデータ整備からやり直したりということを繰り返しまして、最終的にこれでいいだろうと思う値を算出することでモデリングは完了となります。

先ほど紹介したバイブルとして用いた文献1には、主にこの①から③についてのことが書いてあります。

それでは、モデリングの手順を個別に見ていきます。まず、モデルの対象ですが、自動車保険の分析を行っていますので、ここに挙げたとおり、車両や対人、対物などの担保に分けてモデル化を行いました。

①モデル対象の決定(1)

- 担保別にモデル化

- 車両
- 対人賠償
- 対物賠償
- 搭乗者傷害
- 人身傷害
- 代車費用

※担保合計のモデル化は行ってない

9

続いて、担保別のクレームコストのモデル化方法ですが、クレームコストのモデル化には二つの方法があります。一つは、クレーム頻度とクレーム単価をそれぞれモデル化した後、それらの積とする方法です。もう一つが、クレームコストを直接モデル化する方法です。前者の方法の方がより精緻な方法とは思いますが、後者の方法は、作成するモデルが一つのため、モデル作成をより簡便に行えるというメリットがありましたので、後者の方法を採用し、クレームコストを直接モデル化することを行いました。

①モデル対象の決定(2)

- 最終的には、担保別のクレームコストをモデル化したい
- クレームコストのモデル化には、次の2つの方法がある

方法	メリット
クレーム頻度、クレーム単価をそれぞれモデル化後、それらの積とする方法	• 較差の構造を把握しやすい (頻度による較差なのか、単価による較差なのか分かる) • モデルの攪乱要因をより少なくすることができる
クレームコストを直接モデル化する方法	• 作成するモデルが1つのため、モデル作成をより簡便に行える

⇒まずは簡便な方法である、クレームコストを直接モデル化する方法を採択した

10

次に、データ整備についてですが、モデル化用のデータとしては、こちらの表にありますようなレイアウト、つまりは、契約1件別に担保別の経過件数、クレーム件数、クレーム額、それから契約内容を保持させた作りのデータを用意しました。量としては直近3年分を用いました。弊社の場合、このようなデータがシステムの取られているというわけではなく、契約データとクレームデータのデータベース自体が分かれていますので、それらのデータを証券番号をキーとして連結したり、年度をまたぐ契約については、契約レコードを年度で分割したりといったことや、契約の異動があれば、契約レコードを異動前後で分割したりといったことを行って、こういうデータベースを作成しました。

②モデル化用のデータ整備(1)

- 契約1件別に、以下のレイアウトのデータを用意した

- 直近3年分のデータを用いた
- 契約データとクレームデータを証券番号をキーとして連結した
- 年度を跨ぐ契約については、契約レコードを年度で分割した
- 契約の異動があれば、契約レコードを異動前後で分割した

証券番号	担保別に設定			契約内容			...
	経過件数	クレーム件数	クレーム額	観察年度	車種	年齢条件	
00001	0.40	0	0	22	A	All	
00001	0.60	0	0	23	A	All	
00002	0.21	1	10,000	22	A	21	
00002	0.35	0	0	22	B	21	
...							11

②モデル化用のデータ整備(2)

- クレームデータには、一定の処理遅れ期間を考慮するとともに、できる限りクレームを捕捉すべく、普通備金データを活用した
- ダミー変数として、観察年度を追加した(年度によるIBNR不足、攪乱要因を吸収させるため)
- データ項目が未入力の場合は、Unknownとした
- データ項目の入力内容が間違っている場合は、正しい値に振り分けするか、Unknownとした

⇒データの記録のされ方を知る必要あり

モデリングの一連の作業の中で、一つの山場

さらに、クレームデータに関しては、一定の処理遅れ期間を考慮するとともに、できるだけクレームを捕捉したいという観点から普通備金データも活用しました。他には、年度によるIBNR備金の不足や、年度による攪乱要因などを吸収させるために、観察年度をダミー変数として追加しました。さらに、データ項目のクリーニングを行いました。

このように、データ整備では、データを分割したり、統合したり、クリーニングしたりする必要がありますので、結構大変な作業となりました。ですので、モデリングの一連の作業の中でも、データ整備は一つの山場になろうかと思えます。この辺のデータの持ち方についても、先ほど紹介した文献1で結構詳しく紹介されています。

③モデル形状の選択(1)

- 確率分布⇒複合ポアソン-ガンマ分布
定性的な判断のため、他の分布は考慮しなかった
- リンク関数⇒log関数(説明変数の乗算形式)
定性的な判断のため、他の関数は考慮しなかった
- パーセンタイル点によるクレームデータのキャッピング
再現性がないと考えられる高額データをモデルに入れないよう、人系では95%タイル点、対物では99.5%タイル点でキャッピングした

13

続きまして、データ整備が終わりますと、いよいよ本格的にモデリングに突入していきます。今、クレームコストをモデル化しますので、確率分布としては特に悩まずに、複合ポアソン-ガンマ分布を選択しました。リンク関数に関しても、自動車保険の料率が乗算形式でできている実態を鑑み、特に悩まずにlog関数を用いました。

次に、クレームデータのキャッピングを行ったのですが、これはかなり悩んだ部分です。といいますのは、先ほどの文献1に、クレーム額に異常値がないかをチェックしましょうということが書かれているのですが、どの値を異常値と考えるのか、つまりは再現性のない値と考えるのかについては書かれていませんでしたので、具体的にどうしたらよいかがよく分からなかったためです。ですので、試行錯誤を行い、クレーム額を何らかのパーセンタイル点で切ってみるという対応、具体的には、ここに書いてありますように、人系では95%タイル点、対物では99.5%タイル点でクレームデータをキャッピングするといった対応を行いました。これはあくまで1例ですので、この対応方法が絶対によいかは分かりませんが、このような対応を行ったというものです。

続いて説明変数の絞り込みですが、これは、どの項目を説明変数として用いるかということです。モデルには、まずは使えそうな項目をとりあえず全部入れておくのですが、そこから逸脱度テストや、時間一貫性

の傾向があるか、つまりは観察年度によらず同じ傾向を示すかによって、機械的に絞り込み対象を選択していきました。そして、実際に絞り込む際には、極力変数の意味を考えながら、本当にこれを外してしまっていないか、これを入れていいのかということを考えて、絞り込みを行いました。

③モデル形状の選択(2)

- 説明変数の絞り込み
逸脱度テスト(尤度比検定)、時間一貫性(観察年度によらず同じ傾向を示すこと)によって、機械的に絞り込み対象を選択するとともに、変数の意味を考えながら絞り込んだ
 - 説明変数のグルーピング
較差の信頼区間が大きいものを対象に、変数の意味を踏まえつつ行った
 - 交互作用の考慮
現行料率で考慮しているものを中心に考慮
- ⇒説明変数の選択は、つきつめれば、算出結果をどう解釈するかということ

14

次に、説明変数のグルーピングということで、これは説明変数の区分を統合するというもので、例えば年齢条件の26歳以上限定と30歳以上限定を統合するといったものですが、GLMでは較差の信頼区間が計算できますので、信頼区間が大きいものなどを対象に、ここでもできるだけ変数の意味を考えながら統合を行いました。

あとは、やや細かい話ですが、交互作用というものがありまして、これはよく言われるのは、年齢と性別の関係のように、年齢較差が性別で異なるといったものです。交互作用の統計値による機械的な検出方法について、私自身あまり詳しくなかったため、統計値に基づくというよりも、現行料率で考慮しているものを中心に考慮するという対応を行いました。

説明変数の選択においては、統計値だけではなく、その変数で何が説明できるのか、何を説明したいのかといったことを踏まえる必要がどうしても生じてくると思いますので、説明変数の選択は、突き詰めれば算出結果をどう解釈するのと同じであると感じました。

次にパラメーターの計算に移りますが、これはGLM専用のアクチュアリーソフトウェアがありますので、それを用いています。このソフトはワンクリックで計算可能であったり、グラフに優れているために、視覚的に情報をつかむことが容易であったりといったことにメリットがあったと思います。

続きまして、モデルの妥当性検証ですが、一つには、複数の担当者で別々にモデルを作成し、ディスカッションするというを行いました。これによって、説明変数の選択などでどうしても判断が必要な部分について、より納得感の高いモデルへと修正することができたと思います。加えて、単方向から、実績値の較差とモデルの較差を比較することで、モデルが正しいこと、正確に言えば正しそうなことをですが、それを視覚的に確認しました。

④パラメータ計算

- 専用のアクチュアリーソフトウェアを使用した
 - ・ユーザーフレンドリーな作り(ワンクリックで計算可能など)
 - ・グラフに優れており、視覚的に状況を掴むことが容易
- RやSASなどの統計ソフトウェアで計算できる
例:クレーム頻度(Freq=クレーム件数/経過件数)の計算
説明変数X, Y, Z, 経過件数=N

```
R  
glm(Freq~X+Y+Z, poisson(log), weights=N)
```

```
SAS  
proc genmod data= DATA;  
class X Y Z;  
model Freq = X Y Z  
/ dist=poisson link=log;  
weight N;  
run;
```

15

⑤モデルの妥当性検証

- 複数の担当で別々にモデルを作成
 - ・統計値に基づいて機械的に行う部分では差異は生じない
 - ・判断が入る部分では、当然差異が生じた
 - ・このため、どういった考えで判断したかを議論し、より納得感の高いモデル(より定性的な説明が可能なモデル)へと修正した
- 実績値との比較
 - ・単方向から、実績値の較差とモデルの較差を比較し、視覚的にモデルが正しい(正しそうな)ことを確認した

16

以上によって、定性的な説明ができ、かつ実績値にまざまざ当てはまるモデルができたことと判断しました。

最後に、GLMを導入して分かったことについて述べたいと思います。まず良かった点ですが、GLMによりリスクに対する理解が進みました。当然なのですが、単方向からの分析では分からなかった発見が幾つかありました。また、GLMは全項目の影響を一度に計算しますので、単方向の分析を繰り返す従来手法よりも、結果の考察が行いやすかったというメリットがありました。さらには、較差の信頼区間を算出できますので、算出結果の信頼度を踏まえた議論ができたというメリットもありました。

分かったこと(1)

- GLMによりリスクに対する理解が進んだ
 - ・項目の相関に対する発見があった
 - ・ある担保のみ料率計算に用いている項目で、他の担保にも有用なものがあることが分かった
 - ・全項目の影響を一度に計算するため、単方向からの分析を繰り返す従来手法よりも、結果の考察が行いやすかった
 - ・較差の信頼区間を算出できるので、算出結果の信頼度を踏まえた議論ができた

17

分かったこと(2)

- よい分析にはよいデータが必要である
 - ・Unknownの割合が大きい項目は、説明変数としてあまり意味をなさなかった
 - ・モデルの説明は、項目として持っているものに限定される(一般的なデータで補完する方法もあるかもしれないが)
 - ・今持っているデータの項目数や内容(精度)により、モデルの説明力は制限される
- GLMは定量的な評価のみ与える
 - ・ある項目Aでリスク(クレームコスト)が高かった場合、これは、Aだからリスクが高い…？それとも、リスクが高いからA…？
 - ・これがどちらであるかはGLMは語らない

18

次に、GLMを導入する前に、私が勝手にGLMに対して過剰な期待を持っていたと感じた点ですが、一つは、良い分析には良いデータが必要であるということです。といいますのは、モデルの説明は項目として持っているものに限定されますので、今持っているデータの項目数や精度により、モデルの説明力がおのずと制限されるということです。例えば、右利きと左利きで運転のリスクが違うのかを調べたいと思ったとしても、そのようなデータを持っていない限りはGLMを使っても分からないということです。ですので、データの項目数や精度は大変重要であると思いました。

もう一つが、GLMは定量的な評価のみ与えるということです。例えば、あるAという項目でリスクが高

かった場合に、これはAだからリスクが高いのか、それとも、リスクが高いからAなのか、これがどちらであるかは、GLMは教えてくれないということです。ですので、これがどちらであるかは、自分で考える必要があったということが挙げられます。

まとめ

- 自動車保険のプライシングの補助とするために、GLMを用いてクレームコストの分析を行った
- 実務に参考となる文献は少なく、最初のデータ整備も大変だが、計算はアクチュアリーソフトを用いると比較的容易に行える
- モデルの説明力は、用いたデータの精度に制限を受けるため、データは大変重要である
- GLMにより、今まで分からなかったことが定量的に評価できるが、GLMは定量的な評価のみを与えるため、結果の解釈(定性的な理由の検討)が必要となる
- 定量的な評価に加えて、定性的な理由が把握できれば、今何が起きているのか、今後もそれは起こるのか、などが分かってくるため、それを踏まえた対策に着手できる

19

それでは、まとめです。自動車保険のプライシングの補助とするために、GLMを用いてクレームコストの分析を行ってきました。実務に参考となる文献は少なく、また、最初のデータ整備も大変でしたが、計算はアクチュアリーソフトを用いることで比較的容易に行えました。それから、モデルの説明力は、用いたデータの精度に制限を受けますので、データは大変重要であるということに改めて気づかされました。また、GLMにより、今まで分からなかったことが定量的に評価できますが、GLMは定量的な評価のみを与えるので、結果については自分で解釈、つまりは定性的な理由を検討していく必要がありました。ですけれども、GLMによる定量的な評価に加えて、定性的な理由も把握できれば、今何が起きているのか、そして今後もそれは起こるのかなどが分かってきますので、それを踏まえた対策に着手できるのではないかと思います。

私の発表は以上です。

【岩沢】 ありがとうございます。さらに引き続きまして、では、高橋さん、よろしくお願いします。

【高橋】 タワーズワトソンの高橋と申します。よろしくお願いいたします。

私の方からは、同じ「GLMは日本に定着するか」というテーマですが、まずGLMとは何かという話をできるだけかみ砕いて概観した上で、GLMにどのような使い方が考えられるのか、そして、その結果どのようなことが期待できるのかということ、海外の事例も踏まえながらご紹介していければと思います。

まずは「GLMとは」ということですが、「一般化線形モデル」という日本語訳がついています。モデルというと、ボックスの非常に有名な言葉があります。「本質的にいかなるモデルであれ正しくないが、中には役

に立つものがある」というものです。いきなりやや否定的な発言をしましたが、ではGLMは役に立つのかということが、これからのお話になります。

GLM(一般化線形モデル)は モデルである

- "essentially, all models are wrong, but some are useful" («本質的に、いかなるモデルであれ正しくないが、中には役立つものがある») — George E. P. Box
- GLMは「役立つモデル」なのか？

2

そもそもモデルはどのように 役立ち得るのか？

- モデルとは
 - 「問題とする事象(対象や諸関係)を模倣し、類比・単純化したもの。また、事象の構造を抽象して論理的に形式化したもの。ことに後者は、予想・発見の機能を持ち、作業仮説の創出を促すので、科学方法論的に有益。模型。」(三省堂 大辞林 第二版)
- なぜモデル化するのか？ — モデルだからできること・モデルでなければできないこと
 - 実績に存在するノイズ(プロセスエラー=偶然によるばらつき)を除去し、期待値を取り出す
 - 過去を説明すると同時に、将来を予測するために使用できる

3

そもそもモデルというものがどのように役立ちうるのかということ、少し考えてみたいと思います。「モデルとは」ということで、ここでは三省堂大辞林から引用しましたが、「問題とする事象を模倣し、類比・単純化したもの。また、事象の構造を抽象して、論理的に形式化したもの。ことに後者は、予想・発見の機能を持ち、作業仮説の創出を促すので、科学方法論的に有益。模型」と定義されています。一読してもすぐに

は分かりにくい説明かも知れませんが、幾つかキーワードが出ているように思います。「模倣し、単純化したもの」とあるので、必ずしも正しい訳ではないのだろうなという風を読めます。先ほどのボックスの言葉とも呼応するものだと思います。一方で、「抽象して論理的に形式化することによって、予想・発見の機能を持つ。作業仮説の創出を促す」といったところが、恐らく役に立つ可能性を示唆するものだろうと思われます。

もう少し、なぜモデル化するのかということについて考えてみましょう。特にGLMを使うときには実績データに当てはめてみて、その中から期待値を取り出すといったことが目的になって来るわけですが、実績には常にノイズ、つまり偶然によるブレが存在します。したがって、その偶然によるブレの部分を除去することによって、シグナルと呼ばれる期待値の部分を取り出そうというのが、モデルの一つの目標、目的になります。こうやって作られたモデルは、元々は過去のデータに当てはめて作ったモデルですが、将来を予測するためにも使えることが期待されます。「予測モデル」という言葉がよく使われますが、実績から将来を見込むことを可能とする、そのような効果が、モデル化することの一つの意義なのだろうと思います。

GLMとは

- 「一般化線形モデル」...線形回帰モデル(単回帰だけでなく重回帰も含む)を一般化したもの
- 何が「一般化」されるのか？
 - リンク関数...平面を微分可能で単調な関数により曲面に変換できる
 - 例えば線形回帰モデルの加算構造を乗算構造に変換できる
 - エラー関数...応答変数が従うと思われる分布を指数型分布族の中から選択できる、または期待値の大きさに応じた誤差の評価ができる
- 標準的なMinimum-Bias法はすべて、GLMに含まれる(Mildenhall, Stephen, "A Systematic Relationship between Minimum Bias and Generalized Linear Models", Proceedings of the Casualty Actuarial Society, LXXXVI, 1999)

4

さて、一般的なモデルの話から、もう少しGLMの話に入っていこうかと思います。日本語訳しますと「一般化線形モデル」となりますが、名前の通り、線形モデルを一般化したものです。ここでいう線形モデルには、単回帰だけでなく重回帰も含まれます。その意味で、一般化線形モデルも多変量解析ということになります。

では、何が一般化されるのでしょうか。キーワードとしては、「リンク関数」と「エラー関数」という二つのものが出てくるのですが、難しい数学の議論はさておきましょう。まず一つ、線形モデルは、直線であったり平面であったりになる訳ですが、それをリンク関数というものを入れることによって、曲面に変換することが可能になります。別の言い方をしますと、線形回帰モデルというものは加算構造を持っている訳ですが、リンク関数によってこれを乗算構造等に変換することができます。

それからもう一つ、エラー関数というものが出てきます。線形回帰モデルは、暗に正規分布を念頭に置いていたきらいがあります。なぜかといいますと、線形回帰では多くの場合、最小二乗法により当てはめをし

ますが、それは期待値の大きさにかかわらず、ノイズ、ばらつきは均等に出てくることを前提にしています。ところが、われわれが知っているデータでは、例えばクレームの金額、損害額は、高くなるほど、それだけばらつきも大きくということが、当然にあります。一般化線形モデルでは、エラー関数として、応答変数が従うと思われる分布を、指数型分布族の中から選択することで、その部分にある程度対処できます。指数型分布族には、おなじみの分布で言いますと、正規分布に始まり、ポアソン分布やガンマ分布、それから二項分布といったものが含まれます。つまりGLMでは、ポアソン分布に従うもの、ガンマ分布に従うものを、説明変数の掛け算の形で近似できるのではないかとということが、ここで示唆されるわけです。

さらに、「標準的なミニマムバイアス法はすべてGLMに含まれる」とも書きました。ミニマムバイアス法は損保で従来使われてきた手法ですが、それが、実はGLMの文脈でかなりの部分が解釈できます。その意味でGLMは、ミニマムバイアス法を一般化したようなものとしても捉えることができます。

損害保険でのGLMの利用(1): 事故件数、保険金単価

- 多くのリスク較差は経験的に乗算係数として表わされる → リンク関数としてlog関数を選択することで、GLMは多変量の乗算構造を持つ
- 事故件数は一般にポアソン分布に従うものと考えられる → エラー関数としてポアソン分布を選択
- 保険金単価の誤差は期待値の大きさに比例するものと考えられる → エラー関数としてガンマ分布を選択
- 説明変数: リスクファクター
- 結果として得られるものは、乗算型のクレームコスト・モデル → 従前から一般に使われて来た料率と同じ構造
 - 各ファクターのリスク較差は多変量解析による → ダブルカウントの問題を解消

5

今お話ししたことを踏まえて、実際に損害保険でGLMを使うときには、どのようなモデルを仮定するのでしょうか。多くのリスク較差は、経験的に乗算型になっており、乗算係数として表されるというのが、われわれが経験的に知っていることです。GLMではリンク関数をlog関数にすることで、乗算構造になります。また、事故件数は一般にポアソン分布に従うものと考えられます。ポアソン分布は先ほどお話ししましたとおり、指数型分布族に属するものですので、これもGLMの中で処理できることになります。

保険金単価については、誤差が期待値の大きさに比例すると考えられるならば、エラー関数としてガンマ分布を選択するというのが一つの選択になります。このように、事故件数についても保険金単価についても、リンク関数・エラー関数を適切に選ぶことによって、われわれの知っている構造をGLMで表現することができます。

説明変数として使用するものは、一般にリスクファクターとしてあげられるようなものです。例えば、年齢と性別や、新居さんの例ですと、安全装置のある・なしや、車種が、説明変数に該当します。

都合のいいことに、GLMの結果として得られるものは、乗算型のクレームコストのモデルとなります。

つまり、われわれがよく知っている掛け算の料率が得られます。その意味では、なぜわざわざこういう新しい手法を使うのかという話になってしまうかも知れませんが、メリットは多変量解析ができるということです。

新居さんの例にもあったとおり、契約に何か相関がある場合には、ダブルカウントの問題であったり、何か歪んだ見方をしてしまったりすることが起きますが、GLMによりそれを解消できます。今まで使っていた料率構造を、もう少し技術的にいい方法で当てはめることができるというのが、GLMを事故件数や保険金単価で利用する場合のメリット、考え方になろうかと思えます。

損害保険でのGLMの利用(2): 継続率、成約率

- 応答変数(継続率、成約率)は0と1の間の値を取る → リンク関数としてlogit関数を選択することで、応答変数は0と1の間の値を取る
- 継続率、成約率は二項分布に従うものと考えられる → エラー関数として二項分布を選択
- 説明変数: リスクファクターに限らず、契約者の属性、競合他社の保険料との比・差、(継続率の場合)契約の履歴(請求、異動の有無等)、更改前後の保険料増減率や差額等、継続率・成約率に影響すると考えられるあらゆるファクター

6

同じように、継続率や成約率というものも、実際にGLMでモデル化することができます。この場合には、応答変数である継続率や成約率は、0と1の間しか取りませんので、リンク関数としてlogit関数を選択するのが一般的です。そうすることで、応答変数は0と1の間の値を取ることになります。さらに継続率、成約率そのものは二項分布に従うと考えられますので、エラー関数として二項分布を選択します。

説明変数については、リスクファクターは大体含めることになるのでしょうが、それに限らず、契約者の属性や、競合他社の保険料との比・差、つまり他社がどのような保険料を出している中で、自社がどのような保険料を出し、それがどれくらい成約につながったのかをモデルにします。また特に継続率のモデルの場合には、前年度の保険料との比や、あるいは契約の履歴も説明変数に加えます。契約の履歴については、例えば請求や異動は、契約者・被保険者と会社との重要なコンタクトポイントになります。そのようなコンタクトがあったときに、いいサービスを提供できたかどうかという点が、実際にモデルの結果として、表れてくることがあります。継続率や成約率では、そのようなものをすべて含めてモデリングすることになります。

ここまでで、GLMが実務上、かなり定型的に使えるということを見ていただけたと思いますが、実際のモデリングの作業はここからが本番です。すべてのファクターを入れて、モデルを当てはめれば答えは出てきますが、それは大抵の場合、あまり良いモデルとは言えません。モデルというものは、パラメーターを増やせば増やすほど当てはまりが良くなります。パラメーターを減らしていけば当てはまりは悪くなっていき

ますが、一方で、過剰に当てはまりがいいというのは、ノイズまで拾っているという可能性を示唆します。われわれがやりたいのは、ノイズを切り捨ててシグナルだけを取り出すことだとするならば、何でもかんでも当てはまればいいということにはなりません。

よいモデル

- モデルのシンプルさと当てはまりの良さは、トレードオフの関係にある
- パーシモニーの原則、オッカムの剃刀...モデルはシンプルであるほどよい
- 実績データを説明する最善の妥協点を探す

← 平均値のみの
モデル

“最適な”モデル

観測値ごとに
パラメータを1
つ割り振る

← モデルの複雑さ(パラメータの数) →

- 参考となる統計指標: パラメータの信頼区間、カイ二乗検定のp値、AIC等
- 将来予測に使用: 観察年度によりリスク較差は一貫しているか?
- 統計指標に頼らない判断も必要: 得られた結果は、リスク、マーケット、ビジネスへの理解に基づいて、理に合っているか?
- 用途に応じてモデルを使い分けることも

7

一番極端な例をあげましょう。スライド中央の図の右端に、「観測値ごとにパラメーターを一つ割り振る」とありますが、証券番号をファクターにすることを考えてみて下さい。モデルは実績にぴったり当てはまりますけれども、来年使えるものではありません。すべての偶然の要素をシグナルとして拾ってしまっているためです。

したがって、実際のモデリングは、シンプルさと当てはまりの良さとの間で妥協点を探していく作業になります。この際に、一般的に使われる指標としては、パラメーターの信頼区間や、これは先ほど新居さんの資料にも、尤度比検定という言葉が出ていましたが、カイ二乗検定のp値、それからAICといった指標が使われます。技術的なところはあまり深入りしませんが、このようなものはすべて、GLMの文脈の中で出てくるものです。梅田さんの資料の中にもAICは何度か出てきたかと思えますけれども、これらの指標はGLMを当てはめていく中で得られるものです。

さらに、モデルを作っていくうえで考えなければならないことがあります。モデルを将来予測に使用するのであれば、観察年度によって較差がこちらにいたり、こちらにいたりしているようでは、来年どうなるのか、あまり自信が持てません。そのような観点から、これも新居さんの資料にありましたけれども、観察年度によりリスク較差が一貫しているのか、時間一貫性を確認する必要があります。

ただし、統計指標に頼らない判断が常に求められます。得られた結果が、リスクなり、マーケットなり、ビジネスなりに対する理解に基づいて、理にかなっているのか、それが常識にかなっているのか、常に問いかける必要があります。よってモデリングにおいては、統計値で判断できる部分もかなりありますけれども、それだけではなく、そのデータから見える外側の世界に対する理解からもモデルの妥当性を判断する必要があります。

このようないろいろな視点から、モデルを作っていきます。「最適なモデル」と書いていますが、答えは一つではありません。用途に応じてモデルを使い分けるために、最適なモデルの判断を変えることもあります。

モデルをシンプルにするために

- ファクターの取捨選択
- 使用するファクター内での区分のグルーピング
- 説明変数を変量とする → 多項式近似により曲線にあてはめる
- 上記はいずれもGLMの枠組みの中で可能

8

モデルをシンプルにするために、一体どのような手を使うのかですが、ファクターの取捨選択、使用するファクターの中での、区分のグルーピング、それから説明変数を変量にするといったことが行われます。区分のグルーピングについては、例えば区分Aと区分Bにあまり差がないのであれば、そこをまとめてしまいます。説明変数を変量とするというのは、カテゴリ変数を数値変数に変えてしまうということなのですが、そうすることで多項式近似によって曲線に当てはめることができ、ばらつきの小さいスムーズな結果が得られます。これは単純にモデルをシンプルにするというだけではなくて、ノイズを排除するという目的でも使用できます。例えば、年齢の較差が1歳ごとにあちらにいたり、こちらにいたりしているところをスムーズにしたいということは、実務上、求められることかと思えます。これらすべてが、GLMの枠組みの中で可能です。

あとは、技術的・実務的な問題への対処ということで、例えば年齢別のリスク較差が男性と女性で異なるといったケースには、交互作用というものをモデルに入れることで対応ができます。あるいは、ノンフリート等級係数のようなマーケット標準の係数がある場合には、オフセット項というものを入れることによって、その係数を所与の較差として与えて、それに相関する係数で全体を調整することで、結果的に次善のモデルを作成することにはなっていますが、対処することが可能です。これらはいずれもGLMの枠組みの中で可能です。

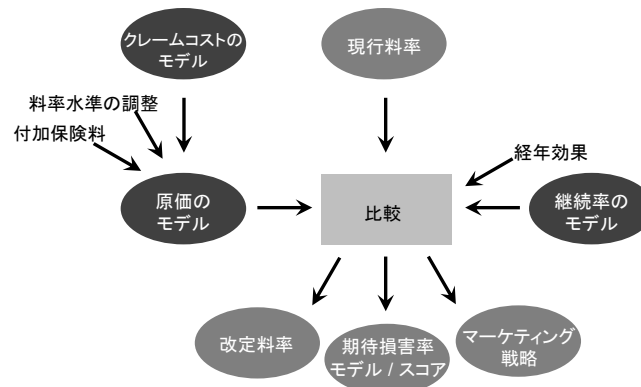
これらですべての問題が解決するとはいいませんが、多くの問題が解決すると思えます。GLMを実際に使ってみると分かりますが、かなり柔軟性があります。完全になんでもできるとはともいいませんけれども、それなりに柔軟にいろいろなことができるということが、ここまで海外で使われている一つの理由なのだろうと考えています。

技術的・実務的な問題への対処

- 例えば、年齢別のリスク較差が男性と女性では異なる場合、年齢×性別の単純な乗算構造ではモデル化できない→年齢別のリスク較差を性別で調整する調整項(交互作用という)を入れることで対処可能
- 例えば、自動車保険におけるノンフリート等級係数のように、マーケット標準の係数があり、それを使用したい場合は、GLMの結果をそのまま使用できない→ノンフリート等級係数は所与の較差として与える(オフセット項という)ことで対処可能(結果として、GLMによる較差と所与の較差との差を、ノンフリート等級と相関のある他のファクターで吸収することで、次善のモデルを作成する)
- 上記はいずれもGLMの枠組みの中で可能

9

得られたモデルをどのように使うか？ — 予測モデルとしての利用



- モデル化することにより、個別の契約に対して、期待値が事前に予測できる
 - 一旦モデル化できれば、契約情報だけでも分析が成り立つ

10

こうして良いモデルが得られたら、これを一体どのように使うのかという利用法、活用法の話になってきます。図の真ん中に「比較」とありますが、何かポートフォリオデータがあるとお考えください。左側にクレームコストのモデル、原価のモデルというものが出てきます。これはGLMで作ったものとお考えください。継続率のモデル、これもGLMで作ったものです。いずれもGLMで作ることが必要な訳ではありませんが、今現在、実務的にGLMが使われているのが実状だと思います。それに対して、例えば現行料率を当ててみると、期待損害率が計算できます。これは契約を取った時点で、計算できます。実績が出るのを待つ必要はありません。モデルがあるので、予測に使えるということが、ここで一つのメリットとなります。

実際に現行料率を当ててみて、それがどう歪んでいるのか分かってくれば、改定料率をどのように作るのかという判断にも使えるでしょう。あるいは、どこで収益性が高いのかが分かれば、マーケティングをどこに注力するのかということに使えるかもしれません。いずれも、モデル化することによって、個別の契約に対して期待値が事前に予測できるということが、このようなモデルを作ることのメリットとなっているものと思います。

損害保険実務における活用例

- 料率改定案の評価(期待損害率、期待継続率、継続契約からの期待収益予測等)
- ポートフォリオの評価(事前の期待損害率・期待収益予測)
- 収益性の高いセグメントの発見 → マーケティング戦略の策定
- 収益の見込めないセグメントの発見 → 引受ルールの策定、選択可能な補償の制限
- 代理店の収益性評価 → 代理店手数料への反映
- 価格感応度 → おすすめパターンによるアップセル

11

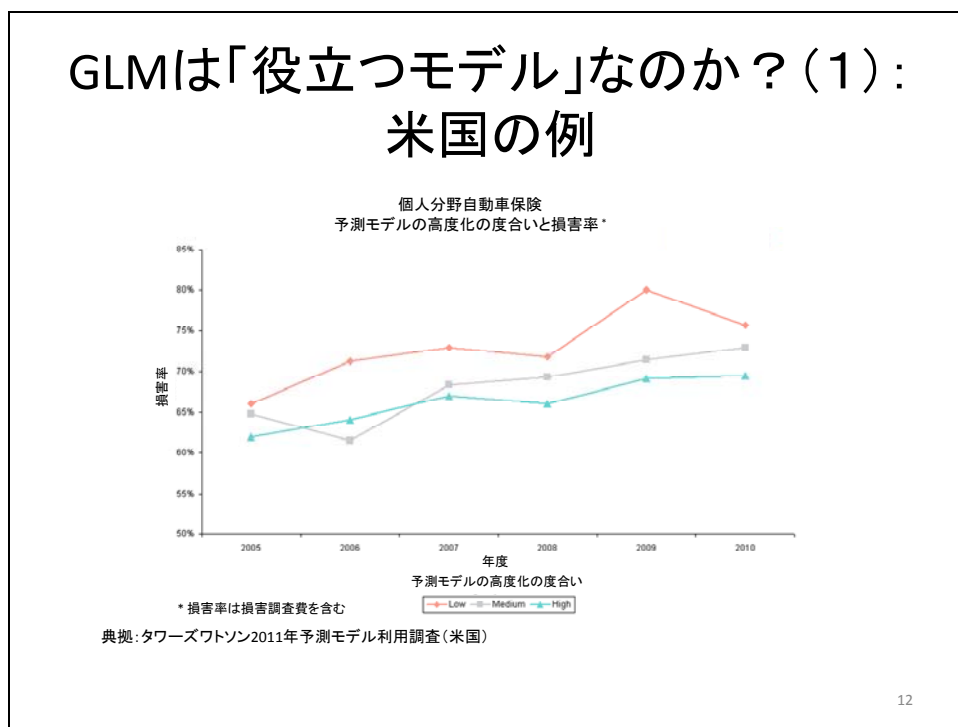
損害保険実務における活用例ですが、実際に海外で見聞きしたのものも含めてあげております。料率改定案の評価については、先ほどお話ししましたとおり、このような料率を世の中に出したら、どれくらいの期待損害率が得られて、継続率がどれくらいになって、そうすると、継続契約からどれくらいの期待収益が得られるのかということが、モデルとしてですが、予測できます。同じように、ポートフォリオがどれくらいの期待損害率を持つのか、どれくらいの期待収益率を持つのかということも予測することができます。あるいは、実際の適用料率がなかなかGLMの結果どおりにできないというときには、収益性の高いセグメント、低いセグメントが出てきてしましますが、そのようなときには、マーケティング戦略、もしくは引き受けルールに反映していくといったことも考えられます。

代理店の収益性評価については、なかなか実務に落とすのは難しいところだと思いますが、中国で実際にやられている例です。代理店ごとに損害率を見た場合、恐らくかなりのノイズを含んでしまうだろうと思います。それをモデル化することによって、ある程度、期待損害率という指標に落とすことができます。それを代理店手数料に反映してしまうというようなアイデアです。このような実務での活用例が、海外には実際にあります。

価格感応度のモデルからお勧めパターンを作って、アップセルに使うといったことも考えられます。これがうまくいくと、日本のマーケットでも活用の可能性があるものと思います。

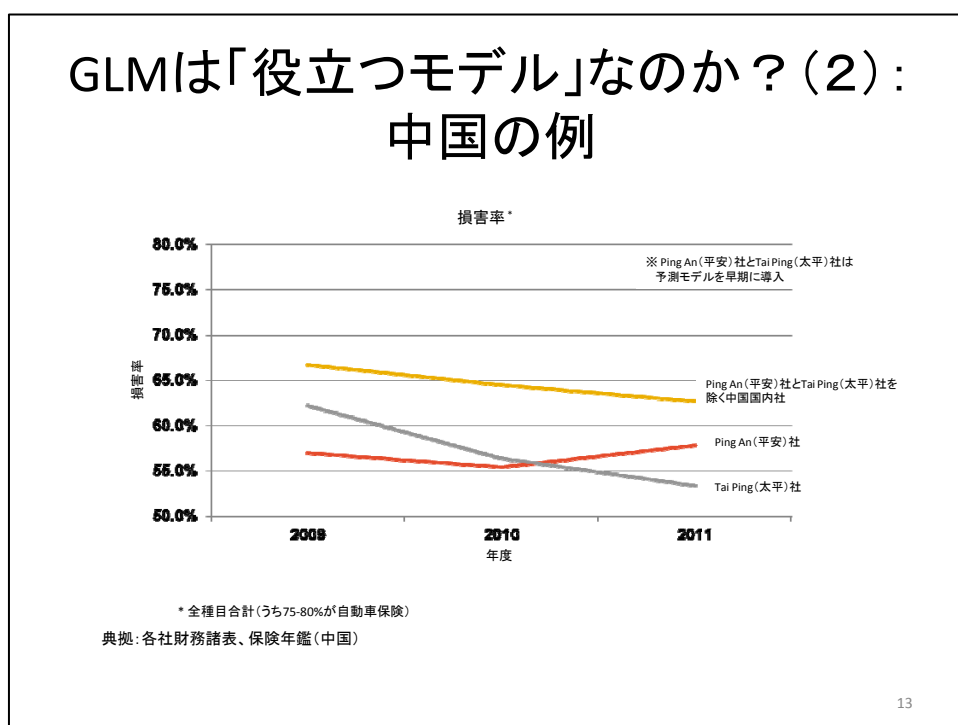
このような海外の活用事例の結果、以下ではGLMが実際に役に立つモデルであったのかという例を三つほど見ていただこうと思います。統計が、アクチュアリーが使うべき厳密さを持っているかという点で、少

し疑問もありますが、ご紹介したいと思います。



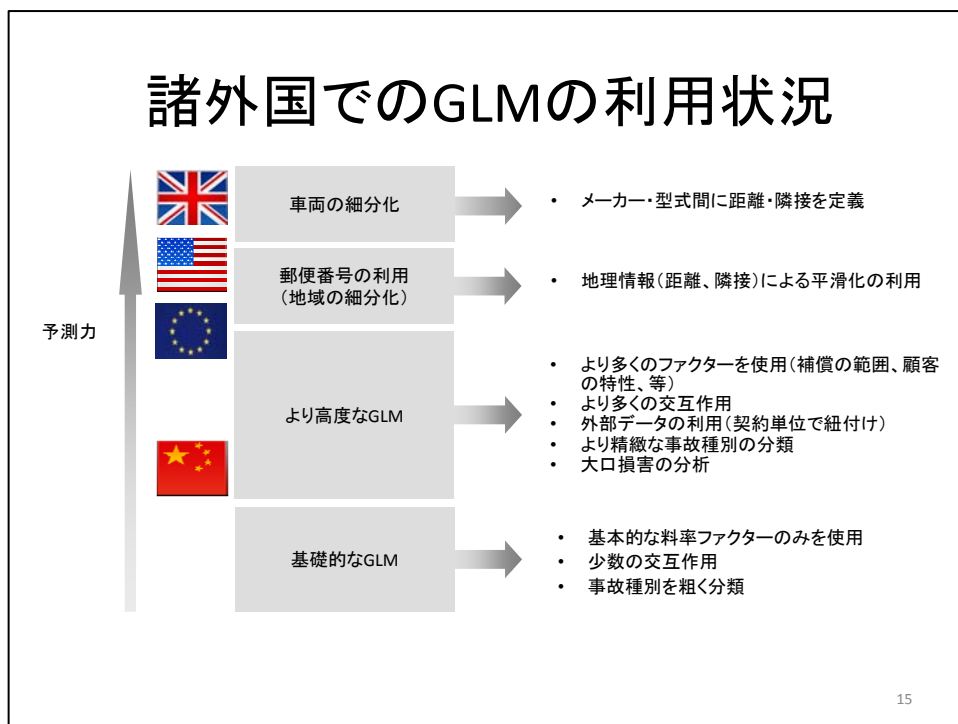
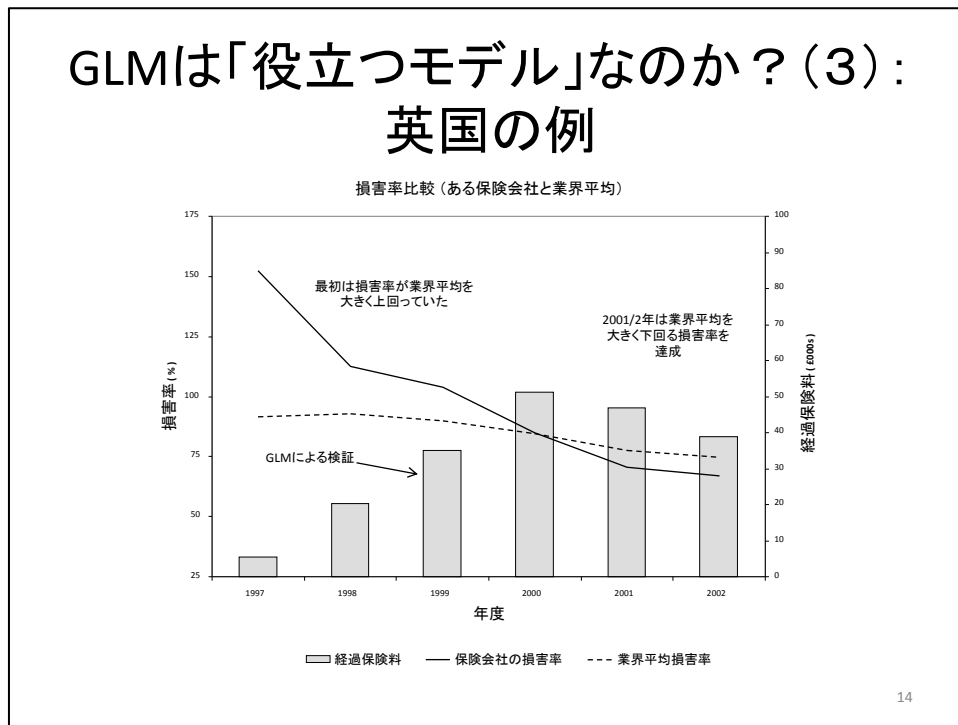
12

これは米国の例です。横軸が年度、縦軸が損害率で、各保険会社を予測モデルの高度化の度合いで低、中、高の三つのグループに分けて、それぞれのグループごとにグラフにしています。高度化の度合いは自己申告ベースなのですが、予測モデルを、われわれは高度なものを使っている、中くらいだ、あまり使っていないと申告したグループに分けて見た場合に、低いと申告したグループの損害率がやはり高く、高度なモデルを使っていると申告したグループの損害率と比べて、2ポイント、3ポイントくらいの差が見えているかと思えます。



13

同じような状況が、中国の例でも見られます。平安社と太平社は、中国で予測モデルを早期に導入した会社なのですが、これら2社は他の会社と比べて損害率が低いのが見てとれます。ここでグラフにしている損害率は、全種目合計ですので、自動車で予測モデルを導入したことが、どのくらい直接この損害率に効いているのか多少の疑問はありますが、それでも際立って損害率が低いという実績が出ています。



3番めにお見せしたいのは英国の例です。これはもう少し直接的な、一つの会社の例です。横軸は年度、棒グラフは経過保険料です。経過保険料が初年度から増えていっているのが、事業をスタートした会社だと

思われます。業界平均の損害率が破線で描かれています。この会社の損害率は、実線で表示されているものですが、元々すごく高かった。ところが、3年めにGLMによる検証を始めた結果、契約量は5年めから少し落ちていますが、損害率はこのような形で、かなり低減することに成功しています。契約量の減少について、実際にどのような判断があったかは分かりませんが、契約を減らしてでも収益性を上げようという判断だったのかも知れません。

海外でのGLMの利用状況の進展について、中国では今、ややGLMがブームのようですが、基礎的なGLMからより高度なGLMへの段階へと進みつつあるようです。大陸ヨーロッパとアメリカが、並んでいるような状況だと思いますが、地域の細分化ということが今やられているようです。これは郵便番号単位での分析です。ただのGLMではないのですけれども、予測モデルとしてはこのようなものが使われ始めています。

英国では、型式による車両の細分化が、郵便番号による地域の細分化と同じようなアイデアで、される段階にきているようです。

アジア圏でも今、中国をはじめ、GLMが使われ始めている状況になってきているものと思います。

それでも「本質的に、いかなるモデルであれ正しくない」

- パラメータエラー／モデルエラーが存在し得る
- 分析の精度には限界がある
- 得られた結果は、リスク、マーケット、ビジネスへの理解に基づいて理に適っているか、再検討が求められる
- 限界を認識した上で使用する必要がある
 - それでもなお、諸外国における活用例は、GLM活用の有用性を示唆

16

ただ、最後にもう一度落とすような話で恐縮なのですが、本質的にいかなるモデルであっても、正しくはありません。使うにあたっては、やはり細心の注意が必要です。そもそも現実の複雑な世界をGLMに当てはめることで、パラメータエラーやモデルエラーが存在し得ます。また分析には、どうしても精度の限界があります。得られた結果についてはリスク、マーケット、ビジネスの理解に基づいて、理にかなっているのかということが常に問いかけとして出てきて、常にジャッジメント、判断が求められます。GLMはこういった限界を認識したうえで、使用する必要がありますが、それでもなお、諸外国でGLMが使われていて、その有用性は先ほど三つの実例をお示ししたとおり、一定期待できるものと考えております。

以上です。

【岩沢】 ありがとうございました。改めまして、お三方、ありがとうございました。素晴らしいパネルを行っていただいたと思います。

このあとディスカッションに入っていきます。ただ、ここでも双方向ツールを使ってみようということで、少し質問させていただきます。皆さんの好みを聞きたいと思っています。具体的な技術的な話よりも一般的な話についてのディスカッションをお聞きになりたいか、もっと具体的な技術的な話をお聞きになりたいか、早いところ質問をさせろというのか、その他、ということで、選んでいただければと思います。

質問5

ディスカッション内容の希望に最も近いものをお選びください。

- A) (個別具体的な話よりも) 少し一般的な点についてのディスカッションを聞きたい
10(25%)
- B) (一般的な話よりも) 個別具体的な点についてのディスカッションを聞きたい
26(65%)
- C) 自分でパネリストたちにいろいろ質問したくてうずうずしている
3(7.5%)
- D) 特に希望はない
1(2.5%)

11

はい、ありがとうございました。結果を見ると、個別具体的な点についてのディスカッションを聞きたいということですので、パネリストの皆さん、ご認識よろしくをお願いします。

一般的な論点:統計学について

- 「統計学は科学であり、技術であり、職人技である」(C. R. ラオ)
- 科学→正しい技術を選択するには専門知識が必要
- 技術→効率的なシステムとして組み込むことが可能
- 職人技→熟練度や経験がものをいう部分がある

12

それでは、出だしは私の方から少し質問をします。これは一般的な論点ですので、会場の要望からしても、私の話は少し端折る感じにしたいと思いますが、それでも、ディスカッションのスタートとしては、このあたりの話から始めたいと思います。会場のみなさんも3人のお話を聞いていろいろお感じになったと思いますが、GLMも、理論の面や技術の面、あるいは、実際に新居さんの話もそうでしたし、高橋さんの話もそうでしたけれども、職人技のような部分もあります。スライドにあるラオはクラメール=ラオの不等式などでよく知られるあのラオです。まだご存命だと思いますけれども、彼は、統計学は科学であって技術であって職人技であると述べています。科学というものは、正しい技術を選択するには専門知識が必要だということでラオは挙げています。技術というものは、効率的なシステムとして組み込むことは可能だとしています。ツールとしてできあがっていますか、そろっていますかという話だと思います。ただ、実際にやろうというときには、かなり判断、経験、その人の技量がいらいます。統計学の場合は統計学者、統計を扱う人の技量がいるし、アクチュアリーだったらアクチュアリーの熟練度や経験がものを言う部分があります。それに対応することをラオも言っています。GLMの話をしていると、そのようなところを意識した方がいいのではないかと思います。

さて、3人のお話を聞いて、日本においてもGLMは、すでにツールとしてといますか、技術としては一応用意されているものはある程度あると理解してよいかと思います。その上で、お三方はいろいろな形でGLMをすでに実践されているわけですが、この科学、技術と職人技のバランスや、習得された順番、どのようにしてGLMを実施してきたか、ご自身の体験や、あるいは周りの人を見てということでも結構なのですけれども、そのあたりについて、コメントをお一人ずついただければと思います。

【高橋】 私自身は、幸か不幸か、科学を勉強する前に、実際にGLMをやれというプレッシャーが大きい環境にいました。そのような意味では、とりあえず技術はそろっている状態で、科学を多少抜きにして、技術からスタートしたようなところがあります。

ただ、個人的な経験から言いますと、その順序は、私にとっては非常によかったと考えています。私は必ずしも数学が得意というわけではありませぬので、実際にデータを見て、何が起きているのだろうということを理解するために、科学を少しずつ学んでいったところがあります。したがって、私は技術からスタートして、職人技は其中で求められて、徐々に身につけた形ですが、科学は後追いになっていました。技術で科学の部分を補っていたと思います。

【新居】 私も高橋さんと大体似たような状況でして、自分がGLMを使うことになったときに有していた統計学の知識は、アクチュアリー試験の数学で問われる程度の内容でしたので、GLM自体何だか分からないという状況でした。ですので、GLMを導入することになって、その担当になってしまったときは、最初はとても不安でした。

その不安は特段解消しないまま、モデリングに突入したのですけれども、実際にモデリングを進めてみての感想としては、高橋さんと大体一緒なのですが、科学や技術の部分は、モデリングを進める中で習得していくのでも、モデリング自体は可能であると思えました。特にアクチュアリーソフトを使う場合においては、なおさら科学・技術の不足ということはあまり問題にならないかと思います。

一方で、実際のデータをどう読み解いたかということは、変数の選択などにダイレクトに影響してきます

し、最終的なモデルの形状にも深く関わってきますので、モデリングにあたっては、データをどう読み解くか、つまりは職人技の部分が最も重要になってくるのかなと思っています。

ただ、これについては、単方向からの従来の分析に携わっていた経験が、実はかなり生きた部分でもありましたので、GLMでなくても、普段からデータ分析に従事されている方にとっては、そのデータに対する読み解き方や理解というものは自然と身につけている部分であると思います。

ということでまとめますと、私としては、GLMに従事する前に科学をしっかり理解することは必ずしも必要ないかと思いますが、不安を減らす程度にはざっと学んでおく方がいいと思います。以上です。

【梅田】 新居さんの答えが完璧すぎて言うことが特にないですが、私自身は試験が終わって大分たってからGLMというものを初めて見ました。知識がありませんでしたので、テキストを探し最初の導入部分だけ理解しました。その後は、アクチュアリーソフトウェアがありましたので、その結果を見ながら考えました。One-way からGLMの係数に移りますと、その時点で係数の変動が非常に大きい場合がありました。特に継続率との関係で、適用料率の設定が非常に難しかったと記憶しています。このような経験からむしろ私の場合は、最後の職人技といえますか、そちらの方が重要だったと思っています。

【岩沢】 ありがとうございます。

さて、先ほどの双方向ツールの結果では、かなり個別具体的な点のディスカッションが希望されているということでしたので、少し時間を見ながら私がコントロールしていきますが、お互いにこのようなことを聞いてみたいということがあれば、質問をし合っていただければと思います。

【梅田】 GLMは信頼区間として、上限から下限まで結果があるのですが、その中でどのポイントを選ぶかというところを苦労していました。そのようなご経験などがありましたら教えてください。

【新居】 私は平均値を選ぶと思います。ただ、信頼区間が広いということは、算出された値の信頼度が低いということだと思いますので、マネジメントから言われたら、平均値の値を伝えますけれども、その値自体の信頼度は低いことも合わせて伝えるかと思います。また、例えば平均値の較差が1を少し超えた1.03くらいだとして、信頼区間を書いてみたら、それが0.9くらいから1.1程度の状況のように、1をまたいだ信頼区間が出ている場合は、結局1.03と出ているでも3%効果があるというのは少し言い過ぎかもしれませんので、そのような場合は、特段影響はないと思いますというような回答をすることが多いかと思います。

いずれにしても、GLMは項目の影響のダブルカウントを排除するために導入されている等、どちらかという極端な値が出にくくなる傾向があるかと思いますので、信頼区間の一番上を言うようなことは、自分はあまりしなかったと思います。

【高橋】 私自身も、今までの経験上はほぼ期待値を使っています。その期待値自体が非常に疑わしい場合、場合によっては元データに戻ったり、個別のデータに戻ったりといったことは行いますが、通常、その期待値よりも信頼区間の方がそれらしいと確信を持ったことはなく、少なくともその信頼区間の中でも最ももつともらしいということが出てきている期待値を、使ってきたのが現状です。ただ、すべてのモデルは間違っていますので、もう一度レビューをするということが求められるのだらうと思います。今、暫定でこの答え

を出したということは、常に答え合わせを求められるものだというのが私の考えです。その意味では、いったん間違っているとしても何でも、答えは出さざるをえません。

そのときに、どの間違いを出すかという意味で、期待値を出しますが、そのあとで、その結果がどれだけ当たっているのか間違っているのかは、常にレビューしなければなりませんし、それでもし、やはりおかしいなということがあったときのために、早急に手当てができる体制を整えておくというのが私の理想です。

【岩沢】 ありがとうございます。他にはどうですか。

【新居】 事故頻度のモデリングに際して、私は基本的にポアソン分布しか使わないのですが、負の二項分布を使った方がいいのではないかとすることが、時々文献などで目にします。実際に負の二項分布を使ってモデル化されるということはよくあるのでしょうか、もしくは、たまにはあるものなのでしょうか。もしあれば、お聞かせいただければと思います。

【梅田】 日常の仕事は、この数年間はI BNRが中心ですので、少しプライシングの方は離れてしまっているのですが、一点でいくらかと問われることはよくあります。分散を広げるということは確かに正しい考え方なのですが、「では、あなたならいくらを選ぶのか」と問われたときに、もっと分散が狭いものがあるといいかなと思うときがあります。実際にはポアソンより分散が狭い分布はありませんので、ポアソンを選んで使っています。

【高橋】 私の方も、非常に限られた経験の中ですし、私自身はほぼ自動車保険の世界でしかGLMは見えていないので、そのような世界での話に限ってですが、事故頻度でポアソン以外を使った、もしくは使っているケースを、私は見たことがありません。ただ、モデリングの最後に、例えば残差のプロットなどを見て、本当に選んだリンク関数、エラー関数が適切かを確認することがあります。つまり、ノイズがノイズらしく均等に出ているかということを確認するのですが、その際に、「これは、本当は」と思ったケースも中にはありました。

ただ、実際には説明のしやすさなども含めて、ポアソンを使ってきました。また、GLMのモデリングでは飽くまで期待値を求めにしているという点で、分散の評価に、過剰に神経質になる必要はないのではないかと考えています。

【岩沢】 ありがとうございます。

残りまだ十数分ありますけれども、いったんフロアの方からご質問があればそれを受け付け、質問があまりなければ再びディスカッションに戻ることになります。

【質問者】 新居さんにお伺いしたいのですが、まず、あなたの説明は大変分かりやすく、内容を理解するのに大変よかったですと思います。

今、お尋ねしたいのは内容のことではなくて、あなたはプレゼンの中で「作業には複数の担当者があつた」と言われていましたね。今のディスカッションの中でも、違った観点からいろいろ論ずるにはそのような方法は面白いと思っているのですが、そこで確認したいのは、何か物事をやるときに、一応意識を統一

するといいますか、あまりばらばらな考え方で各自にやられると、また收拾がつかなくなることもあるでしょうから、この場合は、最初の段階でそのようなことをある程度おやりになってうえて、ある方法に絞って、その中での分担を数人でやったのか、それとも、方法自体を各々の方が自分はこれがいいと思った方法で、まずお互いにやってみて、そのうえて、最後に比較検討をする方法に重点を置かれたのか。

要するに、端的に言えば単なる、単なると言うとおかしいですけども、分担に重点を置いたのか、それとも、各自の方法に重点を置いたのか、その辺を無理に割り切る必要はないのですけれども、その辺をどのような考えでおやりになったか、お伺いしたいです。

【新居】 私が担当していたときは、そのような意味ですと、分担はあまりせず、各自、もう本当に自由に行いました。当初は導入したばかりでしたので、何が正しいかを暗中模索していた状況でもありましたので、そのように行いました。本としてはバイブルとして1冊設定しておりましたので、その本に書いてあることは基本的に踏襲しましたので、確率分布の選択などでは意見の差異は生じませんでした。個々の判断が入りうる部分、主には変数の選択のところで意見の差異が生じました。

当時は、完全に一からGLMを勉強した、私ともう一人の2人で行いましたが、あまり相互に干渉し合うこともなく各々モデルを作り、作った結果をお互いにディスカッションするというのをしばらく繰り返しました。確かに、今となっては一定の方法論が構築できたところもありますので、例えばもう少しパーツごとに分けて分担するというのもできるのかもしれないですが、当時はまだ導入段階で方法が確立していませんでしたので、そのような分担はできなかったというのもあります。

【質問者】 賢明な方法を取られ、結果的にそうなったと私も思います。それで、しかもそれが若い者の教育にもなりますね。大変よかったですと思います。

【新居】 そうですね。ありがとうございます。

【質問者】 損保の経験があまりないことと、GLM自身の知識がないので、あまり貴重な時間を無駄に使いたくないのですが、頭の中にクエスチョンが山のようにたまっています。お聞きしたいことが山のようにあるのですが、このような理解でよろしいのでしょうか。

主に、例えば料率の改定をやったようなときに、その改定がどの程度当たっているのかなど、どちらかという改定そのものよりも、評価のようなところに使い道といいますか、有用性があります、ということでしょうか。場合によっては、マーケティングなどにも活用できる、ひょっとすると午前中の発表にあったERMのようなところにも、何か使い道があるのでしょうか。生保と損保で大分、算方書の書き方なども違いますので、例えばこのようなGLMの結果を料率そのものに反映させようとするとなかなか難しく、従来型の算方書のスタイルに変換する中で何か修正されるのかと思い、お伺いしました。

それと、今、エラー関数のお話もあったのですけれども、例えばリンク関数でも教科書的には山のようにあって、基本的には、損保の料率に使うような場合は、主に、場合によっては加法型もあるのかもしれませんが、基本、乗算型で、一部今はロジットの事例もお勧めいただきましたけれども、あまりそれ以外の凝ったややこしい関数は、解釈的に難しいというのか、まだまだ、むしろ未知の世界があるというように考えた方がいいのかなと思いましたが、いかがでしょうか。

イギリス、アメリカなどの公的なところの第三分野的な商品にGLMを使ったというようなことをお伺いしたことがあるのですが、基本、やはり自動車保険というのが一般的な使われ方なのでしょうか。

あと、今日のお話にはあまりなかったのですが、GLMという手法やソフトそのものに、例えばこのようなことを気をつけた方がいい、など。去年の年次大会などでは、GLMのさらに発展形のようなものもありますというような話があって、ひょっとすると何かまだ改善すべきような要素があって、それに気づいた方が、さらにもう1段先の、GLMでも十分難しいのですが、もっと難しいものを何か開発されているというようなことなのでしょうか。

半分感想で、半分このような理解でよろしいのでしょうかということなのですが、どなたにお聞きしているのかもよく分からないのですが、すみません、質問はそのような感じです。

【高橋】 まず、使い方というお話がありましたが、それは判断ということになるかと思います。実際に、GLMを直接に料率に使えるような状況ならば使うということは、全くない話ではないと思います。ただ、多くの場合、すでにお持ちのポートフォリオがあります。今まで使ってきた保険料があります。そこにいきなりGLMによる料率を使ってしまうと、多分、大激変が起きますが、そういった大激変を起こすのがいい選択であるとは、なかなか思えないケースが多いかと思います。その意味で、やはり評価という形で使うのが、最初はソフトランディングになるのだらうと思います。ただ、海外ではGLMがそのまま料率に使われているケースが多々ありますし、そのような使い方の方が、損害率の安定性は期待できるでしょう。

ここで少し、面白い例があります。収益性という意味では、他社が高い保険料を出しているところで正直に低い保険料を出せば、自社の収益が下がってしまうということがあり得ます。ならば、他社よりも少し安い保険料を出すのが賢いのかも知れません。これは多分に保険料競争の話であって、保険料負担の公平性といった観点は飛ばして話していますが、そのようなケースもありますので、賢い使い方というのが、常にGLMを真正面から料率に使ってしまうことではないということ、海外でもよく話されるものです。

リンク関数に関しては他にいろいろなものがありますが、大体logを使って乗算型にすることが多いでしょう。GLMが使われる領域ということでは、やはり自動車が多いと思いますが、生保ですと、最近では解約率などで使われるようなケースがあります。最後の質問については多分いろいろとありますので、今日お答えするのは難しいかと思います。

【岩沢】 そうですね、もう時間になってしまいました。すみません。またあとで。

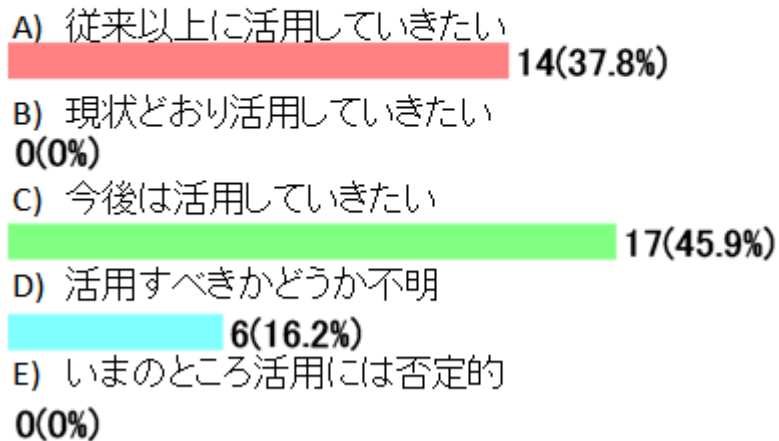
それでは、時間ですので、最後にもう一つだけ双方向ツールでお聞きしたいことがあります。先ほど1度お聞きした質問と同じものをお聞きするもので、われわれとしてはびくびくする質問なのですが、このパネルディスカッションの結果、みなさんのモチベーションが改善されたかどうかお聞きしたいと思います。GLMの今後の利用について、現時点のお考えをお聞かせください。従来以上に活用していきたいのか、現状どおりなのか、今後は活用していきたいのか、活用すべきかどうか不明なのか、今のところ否定的なのかということで、お願いします。

おや、回答総数が減っていますね。これは比較が難しいですね。比較が難しいですが、今後は活用していきたいというのが、見事に増えました。現状どおり活用していきたいという人は、どこにいったのか分からないのですが、2名減っています。ともかくも、先ほどは14、2、14、9、0だったのが14、0、17、6、0となりました。(すでに活用していた人のうち2名が無回答になり、従来活用しておらず「活用すべきか不

明」としていた回答者グループから（ネットで見ても）3名が「今後は活用していきたい」に移ったものと理解される。）ありがとうございました。

質問6（質問4と同じ）

GLMの今後の利用について現時点での考えは下記のうちのどれに近いですか？



それでは、すでに時間が超過してしまいましたが、これでパネルディスカッションを終わります。