ノート

確率のコヒーレント・アグリゲーションの保険数理への応用

西屋知英*　　松山直樹†

2014 年 1 月 31 日投稿
2014 年 2 月 24 日受理

概要

本研究は，Jan Kwiatkowski & Riccardo Rebonato[2011]で提案された確率のコヒーレント・アグリゲーションの手法を，保険数理分野に応用することを目的とする．この手法は，個別の条件付き確率と周辺確率の組み合わせに整合性を持たせるための手法であり，特に，異なる系列の統計データから得られた確率に整合性を持たせることが可能である．
具体的に，複数の公的統計からの副傷病率の推定にこの手法を適用し，その有用性を示す．

キーワード：同時確率，条件付き（周辺）確率，コヒーレント・アグリゲーション，線形計画法，副傷病率

1 はじめに

保険（特に第三分野の保険）に用いられる基礎データは状態数が多いために，一般に個社で収集したデータに基づくことは困難であり，国全体での公的統計資料から様々な確率を推定する必要がある．しかし，公的統計資料は必ずしも保険数理の目的に沿った区分にはなっていないので，複数の公的統計資料を組み合わせることや，データの欠損を補完するアプローチが必要となる．その際に，異なる系列の統計資料間で整合的な確率の観測ができていない可能性がある．このため，保険数理モデルの推定に用いて矛盾する結果を導かないように何らかの手法で元となる観測確率の間の整合性を確保しておく必要がある．

類似の考え方として，例えば，相関行列が外生的に割り当てられているとき，各成分が整合的，すなわち相関行列が意味を持つための条件は，相関行列が半正定値であるということであり，コレスキー分解可能であることを求める場合においては，相関行列が正定値であるということが必要になる．更に，Jan Kwiatkowski & Riccardo Rebonato[2011]では，リスクマネージャーがストレスシナリオを作成する際に，外生的に割り当てた個別の周辺確率と条件付き確率の組み合わせに整合性を持たせ

* 明治大学大学院理工学研究科
† 明治大学総合数理学部　〒164-8525 東京都中野区中野 4-21-1　, ma2yama(at)meiji.ac.jp
本稿作成にあたり，匿名のレフェリーから有益なコメントをいただいた．記して感謝する．

る手法が提案されている．本研究においてもこの論文の提案に習い，整合的な周辺確率と条件付き確率の組み合わせとは，それらの周辺確率と条件付き確率の組み合わせを求めることができるような同時確率が少なくとも一つ存在することであると定義し，その手法の概要を述べるとともに，上述の保険数理分野の問題に適用し，実際の公的統計の数値を用いて手法の有効性を検証する．

本論文の構成は以下のとおりである．2 章では，周辺確率と条件付き確率の組み合わせが整合的であるための条件を導出し，その解法について述べる．3 章では，実際の統計データを用いた適用例を示し，4 章では本手法の課題点について述べる．最後に 5 章で 4 章までの内容についてまとめる．

2 手法の概要

この章では，本研究で用いられるコヒーレント・アグリゲーションの手法について説明する．2.1 節で与えられた周辺確率と条件付き確率の組が整合的であるための制約式を導出し，2.2 節で線形計画法の Phase1 問題の手法を用いることによって，与えられた周辺確率と条件付き確率の組に近い整合的な周辺確率と条件付き確率の組を求める手法を述べる．2.3 節では，2.2 節の線形計画法の制約式の導出に必要となる上下のリミットの設定について述べ，2.4 節で 2.1 節，2.2 節，2.3 節の内容をまとめる．

この章で述べられる手法は，複数条件付き確率に対しても適用することが可能であるが，一般に複数条件付き確率は観測数が少なくなるので統計データから推計することは困難であると考えられるため，周辺確率と 1 条件付き確率に限定して述べることとする．

2.1 制約式の設定

$E_1, E_2, \cdots, E_n$をそれぞれ 0 と 1 の二値のみをとる事象とする．各E_iが二値の事象であることから，その組み合わせに対応する排反な同時事象が2^n個存在するので，それらの同時事象をn次元ベクトル$\boldsymbol{I}_k$（$k = 1,2,\cdots,2^n$)によって表わし，$\boldsymbol{I}_k$全体の集合を $\mathfrak{I}$ で表わす．このとき，各$\boldsymbol{I}_k$に対して同時確率$p[\boldsymbol{I}_k]$を定めることができる．

表 1 n=3 の例

	E_1	E_2	E_3	
$\boldsymbol{I}_1$	0	0	0	$p[\boldsymbol{I}_1]$
$\boldsymbol{I}_2$	0	0	1	$p[\boldsymbol{I}_2]$
$\boldsymbol{I}_3$	0	1	0	$p[\boldsymbol{I}_3]$
$\boldsymbol{I}_4$	0	1	1	$p[\boldsymbol{I}_4]$
$\boldsymbol{I}_5$	1	0	0	$p[\boldsymbol{I}_5]$
$\boldsymbol{I}_6$	1	0	1	$p[\boldsymbol{I}_6]$
$\boldsymbol{I}_7$	1	1	0	$p[\boldsymbol{I}_7]$
$\boldsymbol{I}_8$	1	1	1	$p[\boldsymbol{I}_8]$

以下，$p(E_i=1)$を$\mathrm{P}(E_i)$と表記することとする．

事象$E_1, E_2, \cdots, E_n$に対して，与えられた周辺確率と条件付き確率の組み合わせが整合的であるためには，

$$\sum_{\boldsymbol{I}\in\mathcal{I}} p[\boldsymbol{I}] = \sum_{k=1}^{2^n} p[\boldsymbol{I}_k] = 1 \tag{1}$$

$$0 \le p[\boldsymbol{I}_k] \le 1 \qquad \forall k = 1,2,\cdots,2^n$$

を満たし，各E_iについての周辺確率に対して，

$$\mathrm{P}(E_i) = \sum_{\boldsymbol{I}:\{E_i=1\}} p[\boldsymbol{I}] \tag{2}$$

であり，各E_i, E_jについての条件付き確率に対して，

$$\mathrm{P}(E_i|E_j) = \frac{\mathrm{P}(E_i \cap E_j)}{\mathrm{P}(E_j)} = \frac{\sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}} p[\boldsymbol{I}]}{\sum_{\boldsymbol{I}:\{E_j=1\}} p[\boldsymbol{I}]} \tag{3}$$

であるような，同時確率$p[\boldsymbol{I}_k]$の組が少なくとも一つ存在しなければならない．

本研究ではこのような$p[\boldsymbol{I}_k]$の組が存在しない場合を考え，周辺確率，条件付き確率がそれぞれ，上下のリミット

$$\mathrm{P}^-(E_i) \le \mathrm{P}(E_i) \le \mathrm{P}^+(E_i) \ \ , \ \mathrm{P}^-(E_i|E_j) \le \mathrm{P}(E_i|E_j) \le \mathrm{P}^+(E_i|E_j)$$

の間のどの値をとってもいいとし，その中で整合的な組み合わせを見つける．

このとき，式(2)と上下のリミットの関係から，各事象E_iについて，

$$\mathrm{P}^-(E_i) \le \mathrm{P}(E_i) = \sum_{\boldsymbol{I}:\{E_i=1\}} p[\boldsymbol{I}] \ \Leftrightarrow \ \mathrm{P}^-(E_i) - \sum_{\boldsymbol{I}:\{E_i=1\}} p[\boldsymbol{I}] \le 0 \tag{4}$$

$$\mathrm{P}^+(E_i) \ge \mathrm{P}(E_i) = \sum_{\boldsymbol{I}:\{E_i=1\}} p[\boldsymbol{I}] \ \Leftrightarrow \ \mathrm{P}^+(E_i) - \sum_{\boldsymbol{I}:\{E_i=1\}} p[\boldsymbol{I}] \ge 0 \tag{5}$$

式(3)と上下のリミットの関係から，各事象E_i, E_jについて，

$$\mathrm{P}^-(E_i|E_j) \le \mathrm{P}(E_i|E_j) = \frac{\sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}} p[\boldsymbol{I}]}{\sum_{\boldsymbol{I}:\{E_j=1\}} p[\boldsymbol{I}]} \Leftrightarrow \ \mathrm{P}^-(E_i|E_j) \sum_{\boldsymbol{I}:\{E_j=1\}} p[\boldsymbol{I}] - \sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}} p[\boldsymbol{I}] \le 0 \tag{6}$$

$$\mathrm{P}^+(E_i|E_j) \ge \mathrm{P}(E_i|E_j) = \frac{\sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}} p[\boldsymbol{I}]}{\sum_{\boldsymbol{I}:\{E_j=1\}} p[\boldsymbol{I}]} \Leftrightarrow \ \mathrm{P}^+(E_i|E_j) \sum_{\boldsymbol{I}:\{E_j=1\}} p[\boldsymbol{I}] - \sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}} p[\boldsymbol{I}] \ge 0 \tag{7}$$

が得られ，式(1),(4),(5),(6),(7)を整合的な組み合わせを見つけるための制約式と考える．

2.2 線形計画法による解法

この節では，各事象E_i（$i=1,2,\cdots,n$）に対して与えられた$\mathrm{P}(E_i)$と$\mathrm{P}(E_i|E_j)$の組について，2.1 節で得られた制約式(1),(4),(5),(6),(7)を満たし，そのときの上下のリミットの幅に収まるような$p[\boldsymbol{I}_k]$（$k=1,2,\cdots,2^n$）の組み合わせを求める手法を述べる．

まず，制約式(1),(4),(5),(6),(7)を等式の制約にするために，与えられた各$\mathrm{P}^{\pm}(E_i)$，$\mathrm{P}^{\pm}(E_i|E_j)$に対して，以下のような非負のスラック変数$\pi_i^{\pm}$，$\pi_{i|j}^{\pm}$を定義する．

$$\pi_i^- = -\mathrm{P}^-(E_i) + \sum_{\boldsymbol{I}:\{E_i=1\}} p[\boldsymbol{I}] \tag{8}$$

$$\pi_i^+ = +\mathrm{P}^+(E_i) - \sum_{\boldsymbol{I}:\{E_i=1\}} p[\boldsymbol{I}] \tag{9}$$

$$\pi_{i|j}^- = -\mathrm{P}^-\left(E_i|E_j\right) \sum_{\boldsymbol{I}:\{E_j=1\}} p[\boldsymbol{I}] + \sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}} p[\boldsymbol{I}] \tag{10}$$

$$\pi_{i|j}^+ = +\mathrm{P}^+\left(E_i|E_j\right) \sum_{\boldsymbol{I}:\{E_j=1\}} p[\boldsymbol{I}] - \sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}} p[\boldsymbol{I}] \tag{11}$$

このとき，スラック変数の定め方から，制約式(1),(8),(9),(10),(11)を満たす様な

$$\pi_i^-,\ \pi_i^+,\pi_{i|j}^-,\pi_{i|j}^+,p[\boldsymbol{I}] \geq 0 \quad \forall i,j=1,2,\cdots,n\ ,\ \forall \boldsymbol{I} \in \mathcal{I}$$

が存在すれば，制約式(1),(4),(5),(6),(7)は満たされていることになる．よって，以下では，線形計画法の Phase1 問題の手法を式(1),(8),(9),(10),(11)の制約に応用することにより，式(1),(8),(9),(10),(11)を満たす解を求める．

式(8),(9),(10),(11)はそれぞれ

$$\pi_i^- + \mathrm{P}^-(E_i) - \sum_{\boldsymbol{I}:\{E_i=1\}} p[\boldsymbol{I}] = 0 \tag{12}$$

$$\pi_i^+ - \mathrm{P}^+(E_i) + \sum_{\boldsymbol{I}:\{E_i=1\}} p[\boldsymbol{I}] = 0 \tag{13}$$

$$\pi_{i|j}^- + \mathrm{P}^-\left(E_i|E_j\right) \sum_{\boldsymbol{I}:\{E_j=1\}} p[\boldsymbol{I}] - \sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}} p[\boldsymbol{I}] = 0 \tag{14}$$

$$\pi_{i|j}^+ - \mathrm{P}^+\left(E_i|E_j\right) \sum_{\boldsymbol{I}:\{E_j=1\}} p[\boldsymbol{I}] + \sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}} p[\boldsymbol{I}] = 0 \tag{15}$$

と書き換えることができ，ここで目的関数と制約条件を次に様に書いて

$$w = \sum_i (a_i^- + a_i^+) + \sum_{i\neq j} (a_{i|j}^- + a_{i|j}^+) \tag{16}$$

$$\left.\begin{aligned}
\text{s.t.}\qquad a_i^- &= \pi_i^- + \mathrm{P}^-(E_i) - \sum_{\boldsymbol{I}:\{E_i=1\}} p[\boldsymbol{I}] \\
a_i^+ &= \pi_i^+ - \mathrm{P}^+(E_i) + \sum_{\boldsymbol{I}:\{E_i=1\}} p[\boldsymbol{I}] \\
a_{i|j}^- &= \pi_{i|j}^- + \mathrm{P}^-\left(E_i|E_j\right) \sum_{\boldsymbol{I}:\{E_j=1\}} p[\boldsymbol{I}] - \sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}} p[\boldsymbol{I}] \\
a_{i|j}^+ &= \pi_{i|j}^+ - \mathrm{P}^+\left(E_i|E_j\right) \sum_{\boldsymbol{I}:\{E_j=1\}} p[\boldsymbol{I}] + \sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}} p[\boldsymbol{I}] \\
\sum_{\boldsymbol{I}\in\mathcal{I}} p[\boldsymbol{I}] &= \sum_{k=1}^{2^n} p[\boldsymbol{I}_k] = 1 \\
\pi_i^-,\pi_i^+,\pi_{i|j}^-,\pi_{i|j}^+,a_i^-,a_i^+,a_{i|j}^-,a_{i|j}^+,p[\boldsymbol{I}] &\geq 0 \quad \forall i,j=1,2,\cdots,n,\forall \boldsymbol{I}\in\mathcal{I}
\end{aligned}\right\} \tag{17}$$

線形計画問題（目的関数wを$\mathrm{s.t.}$以下の制約の下で最小化する問題）を解くと，最適解ベクトル$\boldsymbol{x}_{min}$と，その時の目的関数の値$w_{min}(\geq 0)$が求まる（このように置いた線形計画問題は最適解が存在すること

が知られている).

このとき, 各変数が非負であることから$w_{min}=0$であれば$\boldsymbol{x}_{min}$の要素は, a_i^- , a_i^+ , $a_{i|j}^-$, $a_{i|j}^+=0$ $(\forall i,j=1,2,\cdots,n)$となり, (16),(17)の線形計画問題の置き方から, 式(1),(12),(13),(14),(15)を満たしていることが分かり, 同時に(1),(8),(9),(10),(11),の制約式も満たしていることが分かる.

つまり, 最適解ベクトル$\boldsymbol{x}_{min}$を次のように書くとすると

$$\boldsymbol{x}_{min}=(\boldsymbol{x}_{min}^{(1)},\boldsymbol{x}_{min}^{(2)})$$

但し,

$$\boldsymbol{x}_{min}^{(1)}=(p[\boldsymbol{I}_1],\cdots,p[\boldsymbol{I}_{2^n}],\pi_1^-,\cdots,\pi_n^-,\pi_1^+,\cdots,\pi_n^+,\pi_{2|1}^-,\cdots,\pi_{n-1|n}^-,\pi_{2|1}^+,\cdots,\pi_{n-1|n}^+)$$

$$\boldsymbol{x}_{min}^{(2)}=(a_1^-,\cdots,a_n^-,a_1^+,\cdots,a_n^+,a_{2|1}^-,\cdots,a_{n-1|n}^-,a_{2|1}^+,\cdots,a_{n-1|n}^+)$$

$w_{min}=0$であれば, $\boldsymbol{x}_{min}^{(1)}$が制約式(1),(8),(9),(10),(11)の解になっている.

2.3 上下のリミットの設定

次に, 上下のリミットの設定について述べる.

これまでの議論から, 上下のリミットを広げていき, $w_{min}=0$となれば求めたい整合的な周辺確率と条件付き確率の組が求まる, ということが分かる.

上下のリミットの設定にあたって, 最も簡明な方法としては,

(i) $0\le \mathrm{P}^-(E_i),\mathrm{P}^+(E_i),\mathrm{P}^-(E_i|E_j),\mathrm{P}^+(E_i|E_j)\le 1$ であること

(ii) 上下のリミットを広げていく際に, $\mathrm{P}^-(E_i)=\mathrm{P}^+(E_i)=\mathrm{P}(E_i)$, $\mathrm{P}^-(E_i|E_j)=\mathrm{P}^+(E_i|E_j)=\mathrm{P}(E_i|E_j)$から始まり$\mathrm{P}^-(E_i)=\mathrm{P}^-(E_i|E_j)=0$, $\mathrm{P}^+(E_i)=\mathrm{P}^+(E_i|E_j)=1$ で終わるということ

(iii)制御変数が 1 パラメータであること

の 3 条件を充足することが考えられる.

(i)の条件は, この手法を用いて求まった確率が 0 以上 1 以下であることを保証するために必要であり, (ii)の条件は, 本手法をプログラムとして組んだ際にそのプログラムが必ず解を持つことを保証するために必要である. また, 本手法は変数の数が膨大になることから, 制御変数を 1 つにすることが実装上現実的であると考えられる.

このような条件を満たす簡便な手法として Jan Kwiatkowski & Riccardo Rebonato[2011]では以下のような一律のリミットを設定することが提案されている.

$$\mathrm{P}^-(E_i)=(1-\delta)\mathrm{P}(E_i)=\mathrm{P}(E_i)-\delta\mathrm{P}(E_i)$$

$$\mathrm{P}^+(E_i)=(1-\delta)\mathrm{P}(E_i)+\delta=\mathrm{P}(E_i)+\delta\big(1-\mathrm{P}(E_i)\big)$$

$$\mathrm{P}^-(E_i|E_j)=(1-\delta)\mathrm{P}(E_i|E_j)=\mathrm{P}(E_i|E_j)-\delta\mathrm{P}(E_i|E_j)$$

$$\mathrm{P}^+(E_i|E_j)=(1-\delta)\mathrm{P}(E_i|E_j)+\delta=\mathrm{P}(E_i|E_j)+\delta\big(1-\mathrm{P}(E_i|E_j)\big)$$

ここで, $0\le\delta\le 1$とする.

このようにリミットを設定することにより,

$$\mathrm{P}^+(E_i)=(1-\delta)\mathrm{P}(E_i)+\delta=\mathrm{P}(E_i)+\delta\big(1-\mathrm{P}(E_i)\big)\le \mathrm{P}(E_i)+1-\mathrm{P}(E_i)=1$$

であることから, $0\le \mathrm{P}^-(E_i),\mathrm{P}^+(E_i),\mathrm{P}^-(E_i|E_j),\mathrm{P}^+(E_i|E_j)\le 1$となり, (i)の条件を満たしていることが分かる. また, $\delta=0$のとき,

$$\mathrm{P}^-(E_i)=\mathrm{P}(\mathrm{E_i})=\mathrm{P}^+(E_i)\,,\ \mathrm{P}^-\left(E_i\middle|E_j\right)=\mathrm{P}\left(E_i\middle|E_j\right)=\mathrm{P}^+\left(E_i\middle|E_j\right)$$

$\delta=1$のとき，

$$\mathrm{P}^-(E_i)=\mathrm{P}^-\left(E_i\middle|E_j\right)=0\,,\mathrm{P}^+(E_i)=\mathrm{P}^+\left(E_i\middle|E_j\right)=1$$

であるので，δを 0 から 1 まで，例えば1/100刻みで大きくしていくことにより(ii)の条件を満たしていることが分かる．$\delta=0$で$w_{min}=0$となれば，与えられた周辺確率と条件付き確率の組は整合的であったことが分かり，$\delta=1$のときは必ず解が存在し$w_{min}=0$となることが分かる．

したがって，δの初期値を 0 とし，$w_{min}=0$となるまでδを 0 から繰り返し大きくしていくことで，与えられた条件付き（周辺）確率の組に，このリミットの設定の仕方の下で最も近い整合的な条件付き（周辺）確率の組を求めることができる．

2.4 手法のまとめ

(i)　各事象E_i（$\mathrm{i}=1,2,\cdots,\mathrm{n}$）に対して，個別のデータから周辺確率$\mathrm{P}(E_i)$，条件付き確率$\mathrm{P}(E_i|E_j)$を推定する．

(ii)　$\delta=0$とし，上下のリミットを，$\mathrm{P}^-(E_i)=(1-\delta)\mathrm{P}(E_i)$，$\mathrm{P}^+(E_i)=\delta+(1-\delta)\mathrm{P}(E_i)$，$P^-\left(E_i\middle|E_j\right)=(1-\delta)\mathrm{P}\left(E_i\middle|E_j\right)$，$P^+\left(E_i\middle|E_j\right)=\delta+(1-\delta)\mathrm{P}\left(E_i\middle|E_j\right)$のように定める．

(iii)　以下の様な制約式を立てる．

$$\mathrm{P}^-(E_i)-\sum_{\boldsymbol{I}:\{E_i=1\}}p[\boldsymbol{I}]\leq 0$$

$$\mathrm{P}^+(E_i)-\sum_{\boldsymbol{I}:\{E_i=1\}}p[\boldsymbol{I}]\geq 0$$

$$\mathrm{P}^-\left(E_i\middle|E_j\right)\sum_{\boldsymbol{I}:\{E_j=1\}}p[\boldsymbol{I}]-\sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}}p[\boldsymbol{I}]\leq 0$$

$$\mathrm{P}^+\left(E_i\middle|E_j\right)\sum_{\boldsymbol{I}:\{E_j=1\}}p[\boldsymbol{I}]-\sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}}p[\boldsymbol{I}]\geq 0$$

$$\sum_{\boldsymbol{I}\in\mathfrak{I}}p[\boldsymbol{I}]=1$$

$$p[\boldsymbol{I}]\geq 0\ \ \forall\boldsymbol{I}\in\mathfrak{I}$$

(iv)　スラック変数を用いた等式の制約式に変形する．

$$\pi_i^-=-\mathrm{P}^-(E_i)+\sum_{\boldsymbol{I}:\{E_i=1\}}p[\boldsymbol{I}]$$

$$\pi_i^+=\mathrm{P}^+(E_i)-\sum_{\boldsymbol{I}:\{E_i=1\}}p[\boldsymbol{I}]$$

$$\pi_{i|j}^-=-\mathrm{P}^-\left(E_i|E_j\right)\sum_{\boldsymbol{I}:\{E_j=1\}}p[\boldsymbol{I}]+\sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}}p[\boldsymbol{I}]$$

$$\pi_{i|j}^+=\mathrm{P}^+\left(E_i|E_j\right)\sum_{\boldsymbol{I}:\{E_j=1\}}p[\boldsymbol{I}]-\sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}}p[\boldsymbol{I}]$$

$$\sum_{\boldsymbol{I}\in\mathbb{S}} p[\boldsymbol{I}] = 1$$

$$p[\boldsymbol{I}], \pi_i^-, \pi_i^+, \pi_{i|j}^-, \pi_{i|j}^+ \geq 0 \quad \forall\, i, j, \forall \boldsymbol{I} \in \mathbb{S}$$

(v)　以下の様な，(iv)の制約式に付随する線形計画問題を定める．

$$\min \qquad w = \sum_i (a_i^- + a_i^+) + \sum_{i\neq j} (a_{i|j}^- + a_{i|j}^+)$$

$$\text{s.t.} \qquad a_i^- = \pi_i^- + P^-(E_i) - \sum_{\boldsymbol{I}:\{E_i=1\}} p[\boldsymbol{I}]$$

$$a_i^+ = \pi_i^+ - \mathrm{P}^+(E_i) + \sum_{\boldsymbol{I}:\{E_i=1\}} p[\boldsymbol{I}]$$

$$a_{i|j}^- = \pi_{i|j}^- + \mathrm{P}^-\left(E_i|E_j\right) \sum_{\boldsymbol{I}:\{E_j=1\}} p[\boldsymbol{I}] - \sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}} p[\boldsymbol{I}]$$

$$a_{i|j}^+ = \pi_{i|j}^+ - \mathrm{P}^+\left(E_i|E_j\right) \sum_{\boldsymbol{I}:\{E_j=1\}} p[\boldsymbol{I}] + \sum_{\boldsymbol{I}:\{E_i=1,E_j=1\}} p[\boldsymbol{I}]$$

$$\sum_{\boldsymbol{I}\in\mathbb{S}} p[\boldsymbol{I}] = 1$$

$$p[\boldsymbol{I}], \pi_i^-, \pi_i^+, \pi_{i|j}^-, \pi_{i|j}^+, a_i^-, a_i^+, a_{i|j}^-, a_{i|j}^+ \geq 0 \quad \forall\, i, j, \forall \boldsymbol{I} \in \mathbb{S}$$

(vi)　(v)の線形計画問題を解き，その最適解を$\boldsymbol{x}_{min}$とし，その時の wの値をw_{min}とする．

(I) $w_{min} > 0$なら，(ii)に戻りδの値を大きくする．

(II) $w_{min} = 0$なら，最適解

$$\boldsymbol{x}_{min} = (\boldsymbol{x}_{min}^{(1)}, \boldsymbol{x}_{min}^{(2)})$$
$$\boldsymbol{x}_{min}^{(1)} = (p[I_1], \cdots, p[I_{2^n}], \pi_1^-, \cdots, \pi_n^-, \pi_1^+, \cdots, \pi_n^+, \pi_{2|1}^-, \cdots, \pi_{n-1|n}^-, \pi_{2|1}^+, \cdots, \pi_{n-1|n}^+)$$
$$\boldsymbol{x}_{min}^{(2)} = (a_1^-, \cdots, a_n^-, a_1^+, \cdots, a_n^+, a_{2|1}^-, \cdots, a_{n-1|n}^-, a_{2|1}^+, \cdots, a_{n-1|n}^+)$$

の要素である$p[I_1], \cdots, p[I_{2^n}]$から，各$\mathrm{P}(E_i)$，$\mathrm{P}(E_i|E_j)$を求める．このとき，求まった$\mathrm{P}(E_i)$，$\mathrm{P}(E_i|E_j)$の組はこの論文の意味での整合的な組み合わせになっており，(ii)で定めた上下のリミットの間に収まっている．

3 コヒーレント・アグリゲーションの適用例

この章では，2 章で展開したコヒーレント・アグリゲーションの手法を実際の公的統計データに用いることを試みる．具体的に，厚生労働省の患者調査を用いた副傷病率の推定の問題に適用する．

副傷病率を求めるための統計としては，厚生労働省の患者調査の「主傷病×副傷病 推計患者数」があるが，平成 11 年より後の統計では，副傷病として「糖尿病，肥満，高脂血症，高血圧，虚血性心疾患，脳卒中，閉塞性末梢動脈疾患，大動脈疾患，慢性腎不全，精神疾患」しか把握されておらず，がんなどの副傷病率を求める際には平成 11 年以前の統計を用いるしかない．しかし，平成 23 年の統

計では副傷病を複数回答で集計しているが，平成 11 年の統計では副傷病を複数回答で集計していない．これは，平成 11 年の統計では，例えば，がんを主傷病としていて，糖尿病と虚血性心疾患を副傷病としている人は，副傷病として，糖尿病と虚血性心疾患のどちらかしか把握されていないことになる．

そこで，まず 3.1 節では，最新の統計であって，副傷病を複数回答で集計している平成 23 年の統計から推定することのできる確率はそこから求め，それ以外に必要となる確率については平成 11 年の統計から求める．そして，それらが整合的な確率の組み合わせになっていないことを示し，3.2 節でそれらの確率の組に 2 章の手法を適用する．また，3.2 節で線形計画法を解く際には「MATLAB」の「Optimization Toolbox」を使用した．3.3 節でコヒーレント・アグリゲーションを用いて修正された値に対する考察を行う．

3.1 元となる観測確率の算出

この節では，副傷病率（副傷病を持つ確率）を推定するために以下のデータを用いる．

・厚生労働省 平成 23 年 患者調査 閲覧第 98 表

「推計患者数，副傷病 × 主傷病（傷病中分類） × 入院－外来別」

・厚生労働省 平成 11 年 患者調査 閲覧第 45 表

「推計患者数，主傷病のみの者－副傷病を有する者×傷病大分類×入院－外来別」

また，周辺確率を求めるための人口として，総務省の推計人口を用いる．

上でも書いたように，今回は疾病として，がん，糖尿病，虚血性心疾患のみを考えるので条件付き確率と周辺確率の組は以下のようになる．

表 3.1 条件付き確率

P(|)	がん（主）	糖（主）	虚（主）	がん（副）	糖（副）	虚（副）
がん（主）	1	0	0	<u>0.091553</u>	0.101215	0.046694
糖（主）	0	1	0	<u>0.014742</u>	<u>0.044226</u>	0.094664
虚（主）	0	0	1	<u>0.006993</u>	0.206718	<u>0.041958</u>
がん（副）	<u>0.504886</u>	<u>0.019544</u>	<u>0.006515</u>	1		
糖（副）	0.083968	<u>0.037895</u>	0.035826		1	
虚（副）	0.052792	0.067135	<u>0.047619</u>			1

ここで，がん(主)は「がんを主傷病とする事象」，がん(副)は「がんを副傷病とする事象」を意味し，糖尿病，虚血性心疾患についても同様である．また，表の見方としては，それぞれの値が，1 列目に書いてある事象の下での 1 行目に書いてある事象の確率を表している．例えば，P(糖尿病を副傷病とする｜がんを主傷病とする)は第 2 行 6 列であり，その値は 0.101215 である．

表 3.2　周辺確率

	周辺確率
がん(主)	0.002899084
糖(主)	0.001818481
虚(主)	0.000605639
がん(副)	0.000242331
糖(副)	0.00349455
虚(副)	0.000198917

表 3.1，3.2 において，下線の確率は平成 11 年の統計から計算している．

平成 11 年と平成 23 年の統計を用いているので，これらの周辺確率と条件付き確率の組は確率として整合的な組み合わせになっていない．例えば，ベイズの定理から

$$\frac{\mathrm{P}\left(\text{糖}(\text{主})\middle|\text{がん}(\text{副})\right)}{\mathrm{P}\left(\text{がん}(\text{副})\middle|\text{糖}(\text{主})\right)}=\frac{\mathrm{P}\left(\text{糖}(\text{主})\right)}{\mathrm{P}\left(\text{がん}(\text{副})\right)}$$

でなければならないが，

$$\frac{\mathrm{P}\left(\text{糖}(\text{主})\middle|\text{がん}(\text{副})\right)}{\mathrm{P}\left(\text{がん}(\text{副})\middle|\text{糖}(\text{主})\right)}=1.325732879$$

$$\frac{\mathrm{P}\left(\text{糖}(\text{主})\right)}{\mathrm{P}\left(\text{がん}(\text{副})\right)}=7.504120397$$

から分かるように満たしていない．

これは表 3.1，3.2 の確率の組を同時に用いた計算を行った場合，確率として意味を持たない計算をしてしまう危険性を意味している．

3.2 コヒーレント・アグリゲーションによる修正結果

この節では，3.1 節で整合的でないことを示した表 3.1 と表 3.2 の確率の組み合わせを，2 章で紹介したコヒーレント・アグリゲーションの手法を用いて整合的な組み合わせに修正する．

2 章では，すべての確率をそれぞれの上下のリミットの間で修正する方法を示したが，今回の例では，「平成 23 年の統計から得られた確率」は「平成 11 年の統計から得られた確率」よりも信頼できると考え，「平成 23 年の統計から得られた確率」は修正せず，「平成 11 年の統計から得られた確率」のみを修正する．また，糖尿病と虚血性心疾患についての副傷病は平成 23 年度の統計を用いているが，主傷病と比べてデータが少ないことなどから，これらの周辺確率についても修正することとする．

これらを踏まえて，2 章の手法によって計算した結果が表 3.3，3.4 である．

表 3.3　条件付き確率の修正結果

	がん(主)	糖(主)	虚(主)	がん(副)	糖(副)	虚(副)
がん(主)	1	0	0	0.197081405	0.101214575	0.046693657
糖(主)	0	1	0	0.072833445	0.13563578	0.094664358
虚(主)	0	0	1	0.103910017	0.206718213	0.17398776
がん(副)	0.462673417	0.107252572	0.050961165	1		
糖(副)	0.083967756	0.070581637	0.035826243		1	
虚(副)	0.052792188	0.067134574	0.041094481			1

表 3.4　周辺確率の結果

	周辺確率
がん(主)	0.002899084
糖(主)	0.001818481
虚(主)	0.000605639
がん(副)	0.00123490
糖(副)	0.00349455
虚(副)	0.00256418

ここで，下線は修正の対象とした値を表す.

また，これらの周辺確率と条件付き確率は表 3.4 の同時確率から求まる.

表 3.4　修正後の周辺確率と条件付き確率を求めることができる同時確率の組

p(1)	0.9903780609308	p(17)	0.0014213300940	p(33)	0.0021369183760	p(49)	0.0000000000000
p(2)	0.0012078007786	p(18)	0.0000814326344	p(34)	0.0000402928776	p(50)	0.0000000000000
p(3)	0.0018886185908	p(19)	0.0001394116719	p(35)	0.0001221055339	p(51)	0.0000000000000
p(4)	0.0007341491274	p(20)	0.0000438603647	p(36)	0.0000284116638	p(52)	0.0000000000000
p(5)	0.0001517176321	p(21)	0.0000445660194	p(37)	0.0003901349990	p(53)	0.0000000000000
p(6)	0.0001099435184	p(22)	0.0000245011632	p(38)	0.0000383081926	p(54)	0.0000000000000
p(7)	0.0001071037350	p(23)	0.0000410278777	p(39)	0.0001145562573	p(55)	0.0000000000000
p(8)	0.0000994016868	p(24)	0.0000223511747	p(40)	0.0000283561000	p(56)	0.0000000000000
p(9)	0.0004010389396	p(25)	0.0000000000000	p(41)	0.0000000000000	p(57)	0.0000000000000
p(10)	0.0000498604151	p(26)	0.0000000000000	p(42)	0.0000000000000	p(58)	0.0000000000000
p(11)	0.0000622195071	p(27)	0.0000000000000	p(43)	0.0000000000000	p(59)	0.0000000000000
p(12)	0.0000295881791	p(28)	0.0000000000000	p(44)	0.0000000000000	p(60)	0.0000000000000
p(13)	0.0000162506656	p(29)	0.0000000000000	p(45)	0.0000000000000	p(61)	0.0000000000000
p(14)	0.0000132923680	p(30)	0.0000000000000	p(46)	0.0000000000000	p(62)	0.0000000000000
p(15)	0.0000207561148	p(31)	0.0000000000000	p(47)	0.0000000000000	p(63)	0.0000000000000
p(16)	0.0000126328106	p(32)	0.0000000000000	p(48)	0.0000000000000	p(64)	0.0000000000000

3.3 試算結果の考察

以上の結果をまとめると，以下の二点が特徴点として挙げられる．

①平成 11 年のデータから推定した「ある傷病を主傷病とする下でのある傷病を副傷病とする確率（P(副傷病|主傷病)）はすべて妥当な範囲で上昇した

②糖尿病を副傷病とする確率は修正されなかった

①については，平成 11 年のデータにおける副傷病数は複数回答となっていないので，P(副傷病|主傷病)が上昇し，平成 23 年のデータから推定した確率に近づいたことは望ましい方向にあると言える．

②については，今回の数値例では，P(がん(主)|糖尿病(副))，P(がん(主))，P(糖尿病(副)|がん(主))が固定されてしまっているために，P(糖尿病(副))も固定されてしまったと考えられる．

4 コヒーレント・アグリゲーション手法の課題点

今回の試算では上下のリミットの設定方法として，Jan Kwiatkowski & Riccardo Rebonato[2011]の手法を採用し，特に不自然ではない修正結果が得られたが，この方法で常に説得力のある結果が得られるとは限らないことに注意が必要である．

Rebonato らのリミットの設定の仕方は，$\mathrm{P}(E_i)$について，区間$[0,\mathrm{P}(E_i)]$と区間$[\mathrm{P}(E_i),1]$をそれぞれ$\delta:1-\delta$で按分した点を$\mathrm{P}^-(E_i)$，$\mathrm{P}^+(E_i)$と定めていることになる．この方法は簡便で必ず解が存在する点で優れているが，例えば$\mathrm{P}(E_i)$が極めて小さい場合などには，$\mathrm{P}(E_i)-\mathrm{P}^-(E_i)$に比べて$\mathrm{P}^+(E_i)-\mathrm{P}(E_i)$が大きすぎるという問題がある．本研究の趣旨である保険数理への応用ということを考えると，修正された確率が元の確率から大きく離れてしまった場合には，意味づけが難しくなってしまうということが懸念される．この問題に対する一つの解決策として

$$\mathrm{P}_-(E_i)=\mathrm{P}(E_i)-\theta\sqrt{\mathrm{P}(E_i)\big(1-\mathrm{P}(E_i)\big)}$$

$$\mathrm{P}_+(E_i)=\mathrm{P}(E_i)+\theta\sqrt{\mathrm{P}(E_i)\big(1-\mathrm{P}(E_i)\big)}$$

$$\mathrm{P}_-\big(E_i|E_j\big)=\mathrm{P}\big(E_i|E_j\big)-\theta\sqrt{\mathrm{P}\big(E_i|E_j\big)\Big(1-\mathrm{P}\big(E_i|E_j\big)\Big)}$$

$$\mathrm{P}_+\big(E_i|E_j\big)=\mathrm{P}\big(E_i|E_j\big)+\theta\sqrt{\mathrm{P}\big(E_i|E_j\big)\Big(1-\mathrm{P}\big(E_i|E_j\big)\Big)}$$

の様な範囲$[\mathrm{P}_-(E_i),\mathrm{P}_+(E_i)]$，$\big[\mathrm{P}_-\big(E_i|E_j\big),\mathrm{P}_+\big(E_i|E_j\big)\big]$（但し，$0\le\mathrm{P}_-(E_i),\mathrm{P}_+(E_i),\mathrm{P}_-\big(E_i|E_j\big),\mathrm{P}_+\big(E_i|E_j\big)\le 1$を満たすような$\theta$）を定めた上で，区間$[\mathrm{P}_-(E_i),\mathrm{P}(E_i)]$，区間$[\mathrm{P}(E_i),\mathrm{P}_+(E_i)]$をそれぞれ$\delta:1-\delta$で按分した点を$\mathrm{P}^-(E_i)$，$\mathrm{P}^+(E_i)$と定め，区間$\big[\mathrm{P}_-\big(E_i|E_j\big),\mathrm{P}\big(E_i|E_j\big)\big]$，区間$\big[\mathrm{P}\big(E_i|E_j\big),\mathrm{P}_+\big(E_i|E_j\big)\big]$をそれぞれ$\delta:1-\delta$で按分した点を$\mathrm{P}^-\big(E_i|E_j\big)$，$\mathrm{P}^+\big(E_i|E_j\big)$と定めることも考えられる．このように上下のリミットを設定することで，元の確率からの修正幅を制限することができるが，この場合，解が求まらないということも起こりうる．しかし，もしこのリミットの設定によって解が見つからないのであれば「データとして採用しない」ということにも一定の合理性があるのではないかと考えられる．

5 まとめ

本研究では，確率のコヒーレント・アグリゲーションの手法を紹介し保険数理への応用を論じた．保険数理，特に第三分野の保険については商品性が複雑なことや，用いることのできるデータが限られていることから，基礎率作成の素材として必ずしも整合的でない確率の組み合わせを使わなければならない状況が多々あると考えられる．それらを同時に確率として用いる際には，3 章で示した数値例の様に，整合的な確率の組に修正してから使うことが一つの方策になるのではないかと考えられる．また，3 章の数値例では平成 11 年の統計から求めた確率がある程度修正されたが，この手法を用いても値がほとんど修正されなかった場合には，元々の確率の組み合わせが概ね整合的なものであったということが分かる．このことから，異なる統計表から求めた確率が異常な組み合わせになっていないことを確認するためのツールとしても用いることができると考えられる．

また，本手法は統計データから推定した個別の確率の組を，同時に確率として用いることのできる組にするための手法であり，例えば事象間の因果関係などは本手法で修正された後に考慮すべきものであると考えている．なお，この手法を用いて修正された確率の組は，数学的に整合的な組み合わせになっているというだけであり，特に修正幅が大きかった場合には意味づけが難しくなることに注意が必要である．本手法には上下のリミットの設定の仕方によって求まる解が異なるという問題があるが，求まる解が観測値と最も近くなるような上下のリミットの設定方法については今後の研究課題としたい．

参考文献

[1] Jan Kwiatkowski & Riccardo Rebonato[2011],” A Coherent Aggregation Framework for Stress Testing and Scenario Analysis” *Applied Mathematical Finance Volume 18* 139-154

[2] Riccardo Rebonato[2010],”COHERENT STRESS TESTING” WILEY

[3] 今野浩[1987]『線形計画法』日科技連出版社

[4] 山内恒人[2009]『生命保険数学の基礎』東京大学出版会

Coherent Aggregation Method for Actuarial Use

Tomohide Nishiya　Naoki Matsuyama
Meiji University, 4-21-1 Nakano Nakano-ku Tokyo 164-8525, Japan
ma2yama(at)meiji.ac.jp

Abstract

Coherent Aggregation Method for probabilities proposed by Jan Kwiatkowski & Riccardo Rebonato[2011] intends to ensure consistency of a combination of marginal probabilities and conditional probabilities of stress scenarios by adjusting each probabilities automatically. We propose an application of the method for actuarial use to estimate incidental morbidity rates from different series of public statistics.

Key Words:
joint/marginal/conditional probability, coherent aggregation, linear programming, incidental morbidity