支払備金に関する Mack の公式の一般化

斎藤新悟 (Shingo SAITO)
九州大学大学院数理学研究院

概 要

損害保険会社において，既発生の事故に対する未払いの保険金に対応するために積み立てる金額を支払備金という．支払備金を区間推定する 1 つの方法として，Mack は自身のモデルに基づいて支払備金の推定量の平均 2 乗誤差を推定する公式を与えた．この講演では，Mack のモデルを一般化したモデルにおいて，支払備金を含む種々の値の推定量の平均 2 乗誤差を推定する公式を与える．

1　支払備金とチェーンラダー法

1.1　支払備金とは

損害保険業においては，事故が発生してから保険金支払いまでに損害調査などのプロセスが必要になることから，事故発生と保険金支払の間に時間差が生じることが多い．したがって，既発生の事故に対する未払いの保険金の額を推定し，将来支払うために積み立てておく必要がある．このような未払いの保険金の額を**支払備金** (claims reserve) という：

$$（支払備金）=（最終保険金）-（支払保険金）.$$

1.2　ランオフ三角形

支払備金の推定のためには次のようなデータが用いられる：

		経過年数 j				
		1	2	$\cdots$	$n-1$	n
事故年度 i	1	$C_{1,1}$	$C_{1,2}$	$\cdots$	$C_{1,n-1}$	$C_{1,n}$
	2	$C_{2,1}$	$C_{2,2}$	$\cdots$	$C_{2,n-1}$	$(C_{2,n})$
	$\vdots$	$\vdots$	$\vdots$	$\iddots$	$\vdots$	$\vdots$
	$n-1$	$C_{n-1,1}$	$C_{n-1,2}$	$\cdots$	$(C_{n-1,n-1})$	$(C_{n-1,n})$
	n	$C_{n,1}$	$(C_{n,2})$	$\cdots$	$(C_{n,n-1})$	$(C_{n,n})$

ここで，$C_{i,j}$ $(i,j=1,\ldots,n)$ は第 i 年度に発生した事故に対して，j 年経過するまでに（すなわち第 $i+j-1$ 年度末までに）支払った保険金の総額を表す．第 n 年度末の時点では $i+j \leqq n+1$ なる $C_{i,j}$ が既知であり，カッコで囲まれた $i+j \geqq n+2$ なる $C_{i,j}$ が未知である．既知のデータが左上の三角形部分であることから，このようにまとめられたデータを**ランオフ三角形** (run-off triangle) と呼ぶ．

以下簡単のため，各年度の事故は高々 n 年以内に保険金の支払が完了する，すなわち $i = 1, \dots, n$ に対して $C_{i,n}$ は事故年度 i の事故に対する最終保険金であると仮定する．このとき支払備金 R は

$$R = (C_{2,n} - C_{2,n-1}) + (C_{3,n} - C_{3,n-2}) + \cdots + (C_{n,n} - C_{n,1})$$

となる．

1.3 チェーンラダー法

支払備金 R を推定する方法として最も広く用いられているのが**チェーンラダー法** (chain-ladder method) である．まず，$i + j \geqq n + 2$ に対する $C_{i,j}$ の推定量 $\widehat{C}_{i,j}$ を求めれば，R は

$$\widehat{R} = (\widehat{C}_{2,n} - C_{2,n-1}) + (\widehat{C}_{3,n} - C_{3,n-2}) + \cdots + (\widehat{C}_{n,n} - C_{n,1})$$

で推定できることに注意する．

チェーンラダー法においては，経過年数が j から $j + 1$ になる際に累積支払保険金は事故年度によらず定数 f_j 倍となると考える．すなわち理想的には $C_{i,j+1} = C_{i,j} f_j$ $(i = 1, \dots, n,$ $j = 1, \dots, n-1)$ であり，両辺のずれは誤差に起因するとみなす．

各 $j = 1, \dots, n-1$ に対して，$C_{i,j+1}/C_{i,j}$ の値が既知であるのは $i = 1, \dots, n - j$ のときなので f_j を

$$\widehat{f}_j = \frac{C_{1,j+1} + \cdots + C_{n-j,j+1}}{C_{1,j} + \cdots + C_{n-j,j}}$$

で推定し，それを用いて $i + j \geqq n + 2$ のとき

$$\widehat{C}_{i,j} = C_{i,n+1-i} \widehat{f}_{n+1-i} \cdots \widehat{f}_{j-1}$$

で $C_{i,j}$ を推定する．

1.4 問題

以下では，次の 2 つの問題を考える：

- チェーンラダー法はどのような数学的モデルで正当化できるか？ 特に，$\widehat{f}_j$ が類似の推定量

  $$\widehat{f}'_j = \frac{1}{n-j}\left(\frac{C_{1,j+1}}{C_{1,j}} + \cdots + \frac{C_{n-j,j+1}}{C_{n-j,j}}\right)$$

  より優れていることを保証するモデルはどういったものか？
- 支払備金を区間推定するにはどうすればよいか？

ここで，**区間推定** (interval estimation) とは「95%以上の確率で $90 < R < 110$ である」という形の推定であり，このとき区間 $(90, 110)$ を 95%**信頼区間** (confidence interval) と呼ぶ．それに対してチェーンラダー法のような「R は 100 である」という形の推定を**点推定** (point estimation) という．

次節では，これらの問題を考えるために必要な確率論の基礎事項を述べる．

2　確率変数と条件つき期待値

2.1　確率変数

2.1.1　連続型確率変数

確率変数 (random variable) は確率論においては確率空間上の可測関数として定義されるが，本講演では「定まった確率でランダムな値を取る変数」という理解で十分である．$\mathbb{R}^n$ の値を取る確率変数を **$\boldsymbol{n}$ 次元確率変数** (n-dimensional random variable) といい，以下では単に確率変数といえば 1 次元確率変数を表すものとする．n 次元確率変数 X は n 個の確率変数 $X_1, \ldots,$ X_n を並べたものとみなせる：$X = (X_1, \ldots, X_n)$．

簡単のため，本講演で扱う確率変数はすべて連続型であるとする．ここで，確率変数 X が **連続型** (continuous) であるとは，

- 任意の $x \in \mathbb{R}$ に対して $f(x) \geqq 0$,
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$

を満たす関数 $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ が存在して，$a < b$ なる任意の $a, b \in \mathbb{R}$ に対して

$$P(a < X < b) = \int_a^b f(x)\,dx$$

が成立することをいう．この f を X の **確率密度関数** (probability density function) と呼ぶ．

2.1.2　期待値・分散

連続型確率変数 X の **期待値** (expectation) $E[X]$ を

$$E[X] = \int_{-\infty}^{\infty} x f(x)\,dx$$

で定義する．ただし f は X の確率密度関数であり，右辺が定義されるときに $E[X]$ は定義される．

X の **分散** (variance) $V(X)$ を

$$V(X) = E\bigl[(X - E[X])^2\bigr]$$

で定義する（右辺が定義されるときに $V(X)$ は定義される）．

2.2　確率変数の関係

2.2.1　同時確率密度関数・周辺確率密度関数

2 次元確率変数 (X, Y) が **連続型** (continuous) であるとは，

- 任意の $(x, y) \in \mathbb{R}^2$ に対して $f(x, y) \geqq 0$,
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\,dx\,dy = 1$

を満たす関数 $f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ が存在して，$a<b,\ c<d$ なる任意の $a,b,c,d \in \mathbb{R}$ に対して

$$P(a<X<b,\ c<Y<d) = \int_c^d \int_a^b f(x,y)\,dx\,dy$$

が成立することをいう．この f を (X,Y) の**確率密度関数** (probability density function) と呼ぶ．以下 2.2.3 節まで，(X,Y) を連続型 2 次元確率変数とし，f をその確率密度関数とする．

このとき $g\colon \mathbb{R} \longrightarrow \mathbb{R}$ を $g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$ で定義すると

- 任意の $x \in \mathbb{R}$ に対して $g(x) \geqq 0$，
- $\int_{-\infty}^{\infty} g(x)\,dx = \int_{-\infty}^{\infty} \bigl(\int_{-\infty}^{\infty} f(x,y)\,dy\bigr)\,dx = 1$

であり，$a<b$ なる任意の $a,b \in \mathbb{R}$ に対して

$$\begin{aligned} P(a<X<b) &= \lim_{\substack{c\to-\infty\\ d\to\infty}} P(a<X<b,\ c<Y<d) = \lim_{\substack{c\to-\infty\\ d\to\infty}} \int_c^d \int_a^b f(x,y)\,dx\,dy \\ &= \int_{-\infty}^{\infty} \int_a^b f(x,y)\,dx\,dy = \int_a^b g(x)\,dx \end{aligned}$$

が成立するので，g は X の確率密度関数である．

同様に，$h\colon \mathbb{R} \longrightarrow \mathbb{R}$ を $h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$ で定義すると，h は Y の確率密度関数である．f を**同時確率密度関数** (joint probability density function)，$g,\ h$ を**周辺確率密度関数** (marginal probability density function) と呼ぶ．

2.2.2 条件つき確率

$h(y)>0$ であるような $y \in \mathbb{R}$ に対して，$g(\cdot|y)\colon \mathbb{R} \longrightarrow \mathbb{R}$ を

$$g(x|y) = \frac{f(x,y)}{h(y)}$$

で定義すると

- 任意の $x \in \mathbb{R}$ に対して $g(x|y) \geqq 0$，
- $\int_{-\infty}^{\infty} g(x|y)\,dx = \bigl(\int_{-\infty}^{\infty} f(x,y)\,dx\bigr)/h(y) = 1$

となるので $g(\cdot|y)$ は確率密度関数の条件を満たす．$g(\cdot|y)$ を $Y=y$ の下での X の**条件つき確率密度関数** (conditional probability density function) と呼ぶ．この確率密度関数における期待値

$$\int_{-\infty}^{\infty} xg(x|y)\,dx$$

を $E[X \mid Y=y]$ と書き，$Y=y$ の下での X の**条件つき期待値** (conditional expectation) と呼ぶ．この積分に $y=Y$ を代入して得られる確率変数 $\int_{-\infty}^{\infty} xg(x|Y)\,dx$ を $E[X|Y]$ と書き，Y に関する X の**条件つき期待値** (conditional expectation) と呼ぶ．ここで，$h(y)=0$ となる $y \in \mathbb{R}$ に対しては $E[X \mid Y=y]$ は定義されていなかったが，

$$P\bigl(h(Y)=0\bigr) = \int_{\{y\in\mathbb{R}|h(y)=0\}} h(y)\,dy = 0$$

より $E[X|Y]$ は確率 1 で定義されるので問題ない．

条件つき期待値を用いて，Y に関する X の**条件つき分散** (conditional variance) $V(X|Y)$ を

$$V(X|Y) = E\big[(X - E[X|Y])^2 \big| Y\big]$$

で定義する．

なお，**期待値 $\boldsymbol{E[X]}$，分散 $\boldsymbol{V(X)}$ は実数であるが，条件つき期待値 $\boldsymbol{E[X|Y]}$，条件つき分散 $\boldsymbol{V(X|Y)}$ は確率変数であることに注意**．

2.2.3 独立性

X, Y が**独立** (independent) であるとは，任意の $x, y \in \mathbb{R}$ に対して

$$f(x, y) = g(x)h(y)$$

が成立することをいう．これは，$h(y) \neq 0$ なる任意の $y \in \mathbb{R}$ に対して $g(\cdot|y) = g(\cdot)$ が成立することと同値である．

2.2.4 多次元確率変数の関係

今までは 2 つの 1 次元確率変数の関係を見てきたが，いくつかの多次元確率変数の関係も同様である．例えば，m 次元確率変数 X と n 次元確率変数 Y に対して m 次元確率変数 $E[X|Y]$ が定義される．また，連続型 $n_1 + \cdots + n_m$ 次元確率変数 X を n_1 次元確率変数 X_1，…，n_m 次元確率変数 X_m の組と見たとき，確率密度関数をそれぞれ f, g_1, …, g_m とすると，X_1, …, X_m が独立であるとは任意の $x_1 \in \mathbb{R}^{n_1}$, …, $x_m \in \mathbb{R}^{n_m}$ に対して

$$f(x_1, \ldots, x_m) = g_1(x_1) \cdots g_m(x_m)$$

が成立することをいう．

3 Mack の公式とその拡張

3.1 Mack モデル

1.4 節で述べた 2 つの問題を考察するために Mack が考案したモデルについて考える．このモデルでは $C_{i,j}$ $(i, j = 1, \ldots, n)$ を確率変数と考え，次の 3 つを仮定する：

(1) n 個の n 次元確率変数 $(C_{1,1}, \ldots, C_{1,n})$, …, $(C_{n,1}, \ldots, C_{n,n})$ は独立である．

(2) 各 $j = 1, \ldots, n-1$ に対してある正の定数 f_j が存在して，$i = 1, \ldots, n$ に対して次が成立する：

$$E[C_{i,j+1}|C_{i,1}, \ldots, C_{i,j}] = C_{i,j} f_j.$$

(3) 各 $j=1,\ldots,n-1$ に対してある正の定数 v_j が存在して，$i=1,\ldots,n$ に対して次が成立する：

$$V(C_{i,j+1}|C_{i,1},\ldots,C_{i,j}) = C_{i,j}v_j.$$

ここで，$E[C_{i,j+1}|C_{i,1},\ldots,C_{i,j}]$ は $E[C_{i,j+1}|(C_{i,1},\ldots,C_{i,j})]$（$j$ 次元確率変数 $(C_{i,1},\ldots,C_{i,j})$ に関する確率変数 $C_{i,j+1}$ の条件つき期待値）の略記であり，条件つき分散についても同様である．

このような確率モデルにおいて，パラメータ f_j は定数であるが，その推定量

$$\widehat{f}_j = \frac{C_{1,j+1}+\cdots+C_{n-j,j+1}}{C_{1,j}+\cdots+C_{n-j,j}}$$

は確率変数であることに注意.

仮定 (1), (2) を用いると $\widehat{f}_j = (C_{1,j+1}+\cdots+C_{n-j,j+1})/(C_{1,j}+\cdots+C_{n-j,j})$ が f_j の不偏推定量 (unbiased estimator) である，すなわち $E[\widehat{f}_j] = f_j$ が成立することが分かる．これは $\widehat{f}_j$ を推定量として使うことが妥当であることを示しているが，別の推定量

$$\widehat{f}'_j = \frac{1}{n-j}\left(\frac{C_{1,j+1}}{C_{1,j}}+\cdots+\frac{C_{n-j,j+1}}{C_{n-j,j}}\right)$$

も不偏推定量であることが証明できるので，$\widehat{f}_j$ を使う積極的な理由としては不十分である．そこで仮定 (3) も合わせて考えると，$V(\widehat{f}_j) \leqq V(\widehat{f}'_j)$ であることが分かり，さらにある種の推定量の族の中で $\widehat{f}_j$ は分散を最小にすることが示される．不偏推定量の分散が小さいということは，推定量と真の値とのブレが小さいということなので，よい推定量であることを意味する．したがって $\widehat{f}_j$ は Mack モデルにおいて最良の推定量であると考えられる．

3.2　支払備金の区間推定

連続型確率変数 X の確率密度関数 f は期待値 $\mu = E[X]$ と分散 $v = V(X)$ だけでは定まらないが,

$$\begin{aligned}
P(\mu-\sqrt{20v} < X < \mu+\sqrt{20v}\,) &= \int_{\mu-\sqrt{20v}}^{\mu+\sqrt{20v}} f(x)\,dx \\
&= 1-\left(\int_{-\infty}^{\mu+\sqrt{20v}} + \int_{\mu-\sqrt{20v}}^{\infty}\right) f(x)\,dx \\
&\geqq 1-\frac{1}{20v}\left(\int_{-\infty}^{\mu+\sqrt{20v}} + \int_{\mu-\sqrt{20v}}^{\infty}\right)(x-\mu)^2 f(x)\,dx \\
&\geqq 1-\frac{1}{20v}\int_{-\infty}^{\infty}(x-\mu)^2 f(x)\,dx \\
&= 1-\frac{1}{20} = 0.95
\end{aligned}$$

が成立し（Chebyshev の不等式)，さらに多くの場合に

$$P(\mu-3\sqrt{v} < X < \mu+3\sqrt{v}\,) \geqq 0.95$$

が成立することが知られている．

今の状況では，R の平均・分散の代わりに $\widehat{R}$, $E[(\widehat{R}-R)^2]$ を考えるのが自然であるように思えるが，$C_{i,j}$ $(i+j \leqq n+1)$ が既知であることを考慮すると条件つき期待値 $E[(\widehat{R}-R)^2|C_{\mathrm{known}}]$（ただし $C_{\mathrm{known}} = (C_{i,j})_{i+j\leqq n+1}$ は $n(n+1)/2$ 次元確率変数）を考える方がよいことが分かる．この条件つき期待値を $\widehat{R}$ の**平均 2 乗誤差** (mean squared error) といい，$\mathrm{mse}\,\widehat{R}$ と書く．これを用いて例えば

$$\big(\widehat{R} - 3(\mathrm{mse}\,\widehat{R})^{1/2}, \widehat{R} + 3(\mathrm{mse}\,\widehat{R})^{1/2}\big)$$

を R_{total} の 95% 信頼区間とみなすことができる．

Mack は $\mathrm{mse}\,\widehat{R}$ の推定量として次を提示した：

$$\sum_{i=2}^{n}\left(\widehat{C}_{i,n}^2 \sum_{l=n+1-i}^{n-1} \frac{\widehat{v}_l}{\widehat{f}_l^2}\left(\frac{1}{\widehat{C}_{i,l}} + \frac{1}{\sum_{m=1}^{n-l} C_{m,l}}\right)\right) + 2\sum_{i=2}^{n}\widehat{C}_{i,n}\left(\sum_{i'=i+1}^{n}\widehat{C}_{i',n}\right)\left(\sum_{l=n+1-i}^{n-1}\frac{\widehat{v}_l}{\widehat{f}_l^2\sum_{m=1}^{n-l} C_{m,l}}\right).$$

ただし

$$\widehat{v}_j = \frac{1}{n-j-1}\sum_{i=1}^{n-j} C_{i,j}\left(\frac{C_{i,j+1}}{C_{i,j}} - \widehat{f}_j\right)^2$$

は v_j の不偏推定量である．この $\mathrm{mse}\,\widehat{R}$ の推定量を **Mack の公式**と呼ぶ．

3.3 Mack の公式の拡張

この節では本講演の主結果を述べる．この結果は Mack の公式を 2 つの観点から拡張した公式を与えるものである．

まず Mack モデルを拡張したモデルを考える．Mack モデルには 3 つの仮定があったが，仮定 (3) は 2 次の量である $V(C_{i,j+1}|C_{i,1},\ldots,C_{i,j})$ が $C_{i,j}$ に比例するという仮定である．これはチェーンラダー法を正当化するために不可欠なものであったが，他の仮定に比べて必ずしも自然であるとはいえない．そこで α を任意の固定された実数とし，仮定 (3) を次の仮定 (3′) で置き換える：

(3′) 各 $j = 1,\ldots,n-1$ に対してある正の定数 v_j が存在して，$i = 1,\ldots,n$ に対して次が成立する：

$$V(C_{i,j+1}|C_{i,1},\ldots,C_{i,j}) = C_{i,j}^{\alpha} v_j.$$

Mack モデルは $\alpha = 1$ の場合である．このモデルにおいては，Mack モデルの下での推定量 $\widehat{f}_j$, $\widehat{v}_j$ を使うことはできず，次の推定量を用いる必要がある：

$$\widehat{f}_j = \frac{\sum_{i=1}^{n-j} C_{i,j}^{1-\alpha} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}^{2-\alpha}}, \qquad \widehat{v}_j = \frac{1}{n-j-1}\sum_{i=1}^{n-j} C_{i,j}^{2-\alpha}\left(\frac{C_{i,j+1}}{C_{i,j}} - \widehat{f}_j\right)^2$$

次に，平均 2 乗誤差を考える対象を支払備金以外にも広げる．これは例えば次年度支払保険金 $\sum_{i=2}^{n}(C_{i,n+2-i} - C_{i,n+1-i})$ を区間推定するという実務的要請に応えるものである．各 $i = 1,\ldots,n$

に対して自然数 j_i, k_i は $n+1-i \leqq j_i \leqq k_i \leqq n$ を満たすとし，

$$S = \sum_{i=1}^{n} (C_{i,k_i} - C_{i,j_i})$$

と定義する．S は $\widehat{S} = \sum_{i=1}^{n} (\widehat{C}_{i,k_i} - \widehat{C}_{i,j_i})$ で点推定でき，この推定量の平均 2 乗誤差 $\operatorname{mse}\widehat{S} = E\bigl[(\widehat{S}-S)^2 \big| C_{\text{known}}\bigr]$ を次で推定する：

$$\sum_{i,l=1}^{n} \widehat{\varphi}_{i,l}^2 \widehat{A}_{i,l} + 2 \sum_{1 \leqq i < i' \leqq n} \sum_{l=1}^{n} \widehat{\varphi}_{i,l} \widehat{\varphi}_{i',l} \widehat{B}_l.$$

ただし，$i, l = 1, \ldots, n$ に対して

$$\widehat{A}_{i,l} = \frac{\widehat{v}_l}{\widehat{f}_l^2} \left(\frac{1}{\widehat{C}_{i,l}^{2-\alpha}} + \frac{1}{\sum_{m=1}^{n-l} C_{m,l}^{2-\alpha}} \right), \qquad \widehat{B}_l = \frac{\widehat{v}_l}{\widehat{f}_l^2 \sum_{m=1}^{n-l} C_{m,l}^{2-\alpha}},$$

$$\widehat{\varphi}_{i,l} = \begin{cases} \widehat{C}_{i,k_i} - \widehat{C}_{i,j_i} & (n+1-i \leqq l < j_i), \\ \widehat{C}_{i,k_i} & (j_i \leqq l < k_i), \\ 0 & (\text{それ以外}) \end{cases}$$

である．この推定量の公式が本講演の主結果である．

参考文献

[1] T. Mack, *Distribution-free calculation of the standard error of chain ladder reserve estimates*, ASTIN Bulletin, **23** (1993) no. 2, 213–225. `http://www.casact.org/library/astin/vol23no2/213.pdf` からダウンロード可能．

[2] T. Mack, *Measuring the variability of chain ladder reserve estimates*, Casualty Actuarial Society Forum (1994) Spring, vol. 1, 101–182. `http://www.casact.org/pubs/forum/94spforum/94spf101.pdf` からダウンロード可能．

[3] S. Saito, *Generalisation of Mack's formula for claims reserving with arbitrary exponents for the variance assumption*, Journal of Math-for-industry, **1** (2009A), 7–15. `http://gcoe-mi.jp/publish_list/pub_inner/id:4/cid:9` からダウンロード可能．

[4] 斎藤新悟『Mack の公式：支払備金の区間推定』．谷口説男編，プロシーディング「損保数理に現れる確率モデル」—日新火災・九州大学共同研究 2008 年 11 月研究会—，MI レクチャーノート，**13** (2009)．`http://gcoe-mi.jp/publish_list/pub_inner/id:2/cid:10` からダウンロード可能．

[1], [2] が Mack の原論文であり，Mack の公式の導出も含めその内容は [4] で解説されている．主結果については [3] を参照．