損害保険でのパラメータリスク・モデルリスクのBayes推定による評価

近藤 宏樹[1,2]，斎藤 新悟[2]
[1] 日新火災海上保険株式会社，[2] 九州大学
e-mail : hiroki.kondou@nisshinfire.co.jp, ssaito@artsci.kyushu-u.ac.jp

1 はじめに

本講演は，著者らが日新火災海上保険株式会社と九州大学との共同研究に携わる中で得た研究成果 [1] に基づくものである．

保険会社がリスク管理や商品開発を行う上で，将来の損害率（保険料に対する保険金の割合）を推定することは重要な課題である．本講演では，ある保険種目の損害率の Value at Risk（VaR，分位点）を過去の実績損害率のデータから推定する方法について考察する．

2 従来の手法とその問題点

$\boldsymbol{x} = (x_1, \ldots, x_n)$ を過去 n 年分の実績損害率のデータとし，その標本平均を m_x，（不偏でない）標本分散を s_x^2 とする．従来の典型的な手法は，まず損害率が正規分布 $N(\mu, \sigma^2)$ に従うと仮定して，最尤法でパラメータを $\widehat{\mu} = m_x$, $\widehat{\sigma} = s_x$ と推定し, それを用いて損害率の $100\alpha\%$ VaR を

$$m_x + z_\alpha s_x \tag{1}$$

で推定する（z_α は標準正規分布の $100\alpha\%$分位点）ものである．

この手法は損害率のランダム性に起因する**プロセスリスク**は考慮しているものの，パラメータの推定誤差に起因する**パラメータリスク**やモデルの選択の誤りに起因する**モデルリスク**は考慮していないという点で，改良の余地がある．

3 パラメータリスク

この節では，Bayes 推定を用いて，パラメータリスクを考慮した VaR の推定量を求める方法を提示する．以下, やや記号の乱用であるが, 確率変数とその実現値を区別せずに小文字で表し，確率密度関数・確率関数を常に文字 f で表す．

Bayes 推定においては，分析の対象とする確率変数の分布が持つパラメータも確率変数であると考える．ここでは，損害率 x が正規分布 $N(\mu, \tau^{-1})$ に従うとしたときのパラメータ (μ, τ) がパラメータ空間 $\Theta = \mathbb{R} \times \mathbb{R}_{>0}$ 上の確率変数であると考える．条件つき分布 $x|\mu, \tau$ の確率密度関数は次で与えられる（**尤度**）:

$$f(x|\mu, \tau) = \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau(x-\mu)^2}{2}\right).$$

この状況下で, パラメータ空間上にあらかじめ設定した**事前分布** $f(\mu, \tau)$ と標本 $\boldsymbol{x} = (x_1, \ldots, x_n)$ から**事後分布** $f(\mu, \tau|\boldsymbol{x})$ を

$$\begin{aligned} f(\mu, \tau|\boldsymbol{x}) &\propto f(\mu, \tau) f(\boldsymbol{x}|\mu, \tau) \\ &= f(\mu, \tau) \prod_{i=1}^{n} f(x_i|\mu, \tau) \end{aligned}$$

で計算する．今の状況では，事前分布として正規ガンマ分布を用いることが多い．

定義 1 パラメータ $\alpha, \beta, \delta \in \mathbb{R}_{>0}$, $\gamma \in \mathbb{R}$ の**正規ガンマ分布** $\mathrm{NG}(\alpha, \beta, \gamma, \delta)$ とは，確率密度関数が次で与えられる Θ 上の分布である：

$$\begin{aligned} f(\mu, \tau) = &\sqrt{\frac{\delta}{2\pi}} \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha - \frac{1}{2}} \\ &\times \exp\left(-\beta\tau - \frac{\delta\tau(\mu-\gamma)^2}{2}\right). \end{aligned}$$

注意 2 $\mathrm{NG}(\alpha, \beta, \gamma, \delta)$ は，τ の周辺分布がガンマ分布 $\Gamma(\alpha, \beta)$ であり，条件つき分布 $\mu|\tau$ が正規分布 $N\bigl(\gamma, (\delta\tau)^{-1}\bigr)$ であるような分布である．

(μ, τ) の事前分布が $\mathrm{NG}(\alpha, \beta, \gamma, \delta)$ のとき, 事後分布も $\mathrm{NG}(\alpha', \beta', \gamma', \delta')$ の形となる（**共役事前分布**）.

ここでは [2, Section 3.2] に従って，(μ, τ) の事前分布として $f(\mu, \tau) \propto \tau^{-1}$ なる分布を用いる．$\iint_\Theta \tau^{-1}\, d\mu\, d\tau = \infty$ よりこの条件を正確な意味で満たす分布は存在しない（**変則事前分布**）が，形式的に $\mathrm{NG}(-1/2, 0, 0, 0)$ と考えることにより，事後分布は通常の意味での分布 $\mathrm{NG}\bigl((n-1)/2, ns_x^2/2, m_x, n\bigr)$ となる．

将来の損害率 y の分布（**事後予測分布**）は

$$f(y|\boldsymbol{x}) = \iint_\Theta f(y|\mu, \tau) f(\mu, \tau|\boldsymbol{x})\, d\mu\, d\tau$$

で求められ，y の $100\alpha\%$ VaR は

$$m_x + \sqrt{\frac{n+1}{n-1}}\, t_\alpha(n-1) s_x \tag{2}$$

となる．ただし，$t_\alpha(n-1)$ は自由度 $n-1$ の t 分布の $100\alpha\%$分位点である．

4 モデルリスク

この節では [3] を参考に，モデルリスクまで考慮した VaR の推定量を求める方法を提示する．そのために，損害率が正規分布に従うモデル (N) の他に対数正規分布に従うモデル (LN) を考え，$\mathcal{M} = \{\mathrm{N}, \mathrm{LN}\}$ をモデル空間とする．

モデル LN の下での式 (1), (2) の対応物は

$$\exp(m_{\log x} + z_\alpha s_{\log x}), \tag{3}$$

$$\exp\left(m_{\log x} + \sqrt{\frac{n+1}{n-1}}\, t_\alpha(n-1) s_{\log x}\right) \tag{4}$$

となる．ただし $(\log x_1, \ldots, \log x_n)$ の標本平均，標本分散を $m_{\log x}$, $s^2_{\log x}$ と書いた．

モデルの事後分布 $f(\mathrm{N}|\boldsymbol{x})$, $f(\mathrm{LN}|\boldsymbol{x})$ を求めるには，$M \in \mathcal{M}$ に対して

$$f(\boldsymbol{x}|M) = \iint_\Theta f(\boldsymbol{x}|M, \mu, \tau) f(\mu, \tau|M)\, d\mu\, d\tau$$

を計算する必要があるが，Θ 上の事前分布として変則分布を用いたため $f(\mu, \tau|M)$ は well-defined ではない．そこで，事前分布として変則でない分布 $\mathrm{NG}(\alpha, \beta, \gamma, \delta)$ を N, LN の両方で用いる場合の $f(\boldsymbol{x}|\mathrm{N})/f(\boldsymbol{x}|\mathrm{LN})$ を計算すると，$\alpha = -1/2$, $\beta = \gamma = \delta = 0$ を代入しても意味を持つ式が得られ，代入した結果は $f(\boldsymbol{x}|\mathrm{N})/f(\boldsymbol{x}|\mathrm{LN}) = p/(1-p)$ となる．ただし

$$p = \frac{s^{n-1}_{\log x} \prod_{i=1}^n x_i}{s^{n-1}_{\log x} \prod_{i=1}^n x_i + s^{n-1}_x}.$$

したがって，$\mathcal{M}$ 上の事前分布として一様分布 $f(\mathrm{N}) = f(\mathrm{LN}) = 1/2$ を取ると $f(\mathrm{N}|\boldsymbol{x}) = p$, $f(\mathrm{LN}|\boldsymbol{x}) = 1 - p$ となるので，事後予測分布は

$$f(y|\boldsymbol{x}) = p f(y|\boldsymbol{x}, \mathrm{N}) + (1-p) f(y|\boldsymbol{x}, \mathrm{LN})$$

となり，$100\alpha\%$ VaR の推定量 q は，

$$a_q = \frac{q - m_x}{\sqrt{(n+1)/(n-1)}\, s_x},$$

$$b_q = \frac{\log q - m_{\log x}}{\sqrt{(n+1)/(n-1)}\, s_{\log x}}$$

とおいたときに

$$p F_{t(n-1)}(a_q) + (1-p) F_{t(n-1)}(b_q) = \alpha \tag{5}$$

を満たす数として与えられる．ただし $F_{t(n-1)}$ は自由度 $n-1$ の t 分布の分布関数である．

5 数値例

元データ $\boldsymbol{x}$ として

i	1	2	3	4	5
x_i	0.33	0.42	0.37	0.29	0.31
i	6	7	8	9	10
x_i	0.35	0.42	0.29	0.23	0.27

を用いる $(n = 10)$ と，式 (1)–(5) で与えられる 99% VaR の推定値は次の通りである：

	(1)	(2)	(3)	(4)	(5)
分布	N	N	LN	LN	N/LN
プロ	○	○	○	○	○
パラ	×	○	×	○	○
モデ	×	×	×	×	○
VaR	0.466	0.513	0.494	0.571	0.558

ここで，「プロ」・「パラ」・「モデ」はそれぞれプロセスリスク・パラメータリスク・モデルリスクを考慮しているかどうかを示している．

6 まとめ

本講演で想定した状況では，パラメータリスク・モデルリスクを含んだリスク計量が比較的簡単な手続きで行えることが分かった．しかし，正規分布・対数正規分布以外の分布を利用する場合は，パラメータリスクは求められることが多いものの，ここで提示したアプローチではモデルの事後確率が計算できないためモデルリスクは求めることができず，手法の改良が必要となる．

参考文献

[1] H. Kondo and S. Saito, *Bayesian approach to measuring parameter and model risk in loss ratio estimation*, J. Math-for-Ind. **4** (2012), 85–89.

[2] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis*, second ed., Texts in Statistical Science Series, Chapman & Hall/CRC, Boca Raton, FL, 2004.

[3] A. J. G. Cairns, *A discussion of parameter and model uncertainty in insurance*, Insurance Math. Econom. **27** (2000), 313–330.