数理解析研究所講究録 第2114巻 2019年 171-173

マルコフ決定過程におけるリスク解析

名古屋市立大学大学院芸術工学研究科　中国清華大学
Masayuki Kageyama 影山正幸 *
Graduate School of Design and Architecture,
Nagoya City University
Tsinghua University

1　背景と準備

MDPs(Markov decision processes) 理論の初期の 1980 年代までの研究は, ほとんどは最適方程式とその解をみつけるための policy iteration と value iteration に関するものであった [5]. これらの研究は利得の期待値について議論している. ここでは, MDPs において, Conditional Value at Risk をリスク指標として用いることにする.

ボレル集合 X に対して, B_X を X のボレル部分集合の $\sigma-$algebla とする. 任意のボレル集合 X に対して, X 上のすべての有界なボレル可測関数の集合を $B(X)$ と記す. ボレル集合 X と Y に対して, $P(X), P(X|Y)$ を, それぞれ, X 上のすべての確率測度の集合, Y が与えられた時のすべての条件つき確率測度の集合とする. I をある確率空間 $(\Omega, \mathcal{B}, P)$ 上の利得確率変数とする.
$\mathbb{R}$ を実数の集合とする.

定義 1.1（Artzner et al.[1]） 確率変数 $Y_1, Y_2 \in L^1$ に対して, $\mathcal{R} : L^1 \to \overline{\mathbb{R}}$ が次の 4 つの性質を満たすとき, $\mathcal{R}$ は coherent 性を満たすという.

1. (Monotonicity) $\mathcal{R}(Y_1) \leq \mathcal{R}(Y_2)$ whenever $Y_1 \geq Y_2$ a.s.,
2. (Translation Equivariance) $\mathcal{R}(Y + c) = \mathcal{R}(Y) - c$, if $c \in \mathbb{R}$
3. (Positive homogeneity) $\mathcal{R}(\lambda Y) = \lambda \mathcal{R}(Y)$ if $\lambda > 0$,
4. (Convexity) $\mathcal{R}((1-\lambda)Y_0 + \lambda Y_1) \leq (1-\lambda)\mathcal{R}(Y_0) + \lambda \mathcal{R}(Y_1)$ for $0 \leq \lambda \leq 1$.

これらの公理の必然性については, [1] を参照されたい.

* kageyama@sda.nagoya-cu.ac.jp

定義 1.2（Artzner et al.[1]）

$$V@R_\gamma(I) := \inf\{x \in \mathbb{R} | F_{-I}(x) \geq \gamma\} \ (0 \leq \gamma \leq 1),$$
$$CV@R_\gamma(I) := \frac{1}{1-\gamma}\int_\gamma^1 V@R_p(I)dp \ (0 \leq \gamma \leq 1). \tag{1}$$

ただし, $F_I(x) := P(I \leq x)(x \in \mathbb{R})$.

定理 1.1（Artzner et al.[1]） $CV@R_\gamma(I)$ は coherent 性を満たす.

定理 1.2（Rocafellar et al.[9]）

$$CV@R_\gamma(I) = \inf_{b\in\mathbb{R}}\left\{b + \frac{1}{1-\gamma}E[[-I-b]^+]\right\}. \tag{2}$$

ただし, $[x]^+ := \max\{x, 0\}$.

ボレル集合 $\mathcal{S}, \mathcal{A}$ をそれぞれ, state space, action space とする. $A(x)$ をシステムが x にいる状態の時の実行可能な action の集合とする. $\mathcal{Q} \in P(\mathcal{S}|\mathcal{SA})$ を推移法則, $\tilde{r} \in B(\mathcal{SAS})$ を immediate reward, ν を初期分布とする. X_t, Δ_t を時刻 $t(t \geq 0)$ における状態と action とする. $\prod$ をすべての policy の集合, つまり, $\pi = (\pi_0, \pi_1, \cdots) \in \prod$ に対して, $\pi_t \in P(\mathcal{A}|\mathcal{S}(\mathcal{AS})^t)$ は, すべての $(x_0, a_0, \cdots, a_{t-1}, x_t) \in \mathcal{S}(\mathcal{AS})^t$ に対して,

$$\pi_t(\mathcal{A}(x_t)|x_0, a_0, \cdots, a_{t-1}, x_t) = 1$$

を満たすものとする.

定義 1.3（Kageyama et al.[6]）

$$\rho_{DS}(\tilde{r}|\pi) := \frac{1}{1-\beta}\sum_{t=1}^{\infty} E_\pi[CV@R_\gamma(\tilde{r}(X_{t-1}, \Delta_{t-1}, X_t)|H_{t-1}].$$

定義 1.4（Kageyama et al.[6]）

$$\rho_{AV}(\tilde{r}|\pi) := \limsup_{T\to\infty}\frac{1}{T}E_\pi[CV@R_\gamma(\tilde{r}(X_{t-1}, \Delta_{t-1}, X_t)|H_{t-1}].$$

Discounted case と Average case の value function を

$$\rho_{DS}(\tilde{r}) := \inf_{\pi\in\prod}\rho_{DS}(\tilde{r}|\pi),$$
$$\rho_{AV}(\tilde{r}) := \inf_{\pi\in\prod}\rho_{AV}(\tilde{r}|\pi)$$

とする.

2 MDPs におけるリスク評価

定理 2.1（Kageyama et al.[6]） 任意の $\pi \in \prod$ に対して, ρ_{DS} と ρ_{AV} は coherent 性を満たす.

任意の $\tilde{r} \in B(\mathcal{SAS})$ に対して,

$$r(x,a) := D^{-1}_{-\tilde{r}}(\gamma|x,a) + \frac{1}{1-\gamma}\int [-\tilde{r}(x,a,y) - D^{-1}_{\tilde{r}}(\gamma|x,a)]^+ Q(dy|x,a)$$

とおく.

定理 2.2 (Kageyama et al.[6]) ある仮定の下 [6] で value function ρ_{DS} は,

$$\rho_{DS}(\tilde{r}) = \int h_{DS}(\tilde{r}|x)\nu(dx)$$

で与えられる. ただし, $h_{DS}(\tilde{r}|\cdot)$ は, 以下の最適方程式の unique な解である.

$$h_{DS}(\tilde{r}|x) = \min_{a\in\mathcal{A}}\{r(x,a) + \beta\int h_{DS}(\tilde{r}|y)Q(dy|x,a)\}$$

for $x \in \mathcal{S}$.

定理 2.3 (Kageyama et al.[6]) ある仮定の下 [6] で, 次式をみたす $\nu \in B(\mathcal{S})$ が存在する.

$$\rho_{AV}(\tilde{r}) + \nu(x) = \min_{a\in\mathcal{A}}\{r(x,a) + \int \nu(y)Q(dy|x,a)\}.$$

参考文献

[1] P. Artzner, F. Delbaen, J. M. Eber, D. Heath and H. Ku, "Coherent measures of risk", *Math. Finance*, 9, 1999, 203-228.

[2] P. Artzner, F. Delbaen, J. M. Eber, D. Heath and H. Ku, "Coherent multiperiod risk adjusted values and Bellman's principle", Ann., Oper. Res., 2007, 152, 5-22.

[3] N. Bauerle, A. Popp, "Risk-sensitive stopping problems for continuous-time markov chains", to appear in *Stochastics*.

[4] N. Bauerle, U. Rieder, "Partially observable risk-sensitive Markov Decision Processes", *Mathematics of Operations Research*, 2017, 42 (4), 1180-1196.

[5] A. Feinberg and A. Shwarltz edited, Handbook of Markov decision processes: Methods and Applications, Kluwer, 2002.

[6] M. Kageyama, T. Fujii, K. Kanefuji and H. Tsubaki, "Conditional Value-at-Risk for Random Immediate Reward Variables in Markov Decision Processes", *American Journal of Computational Mathematics*, 2011, 1, 183-188.

[7] S. Kusuoka, "On law invariant coherent risk measures", *Advances in Mathematical Economics*, Vol.3, Springer, Tokyo, (2001), 83-95.

[8] G. Ch. Phlug, A. Pichler, Multistage Stochastic Optimization, Springer, 2014.

[9] R. T. Rockafellar and S. Uryasev, "Optimization of Conditional Value-at-Risk," *Journal of Risk*, Vol. 2, N 2000, 3, 21-42.