数理解析研究所講究録
947巻 1996年 22-30

A utility deviation in discounted Markov decision processes with general utility

和歌山大教育 門田良信 (Y. Kadota)
千葉大教育 蔵野正美 (M. Kurano)
千葉大 理 安田正実 (M. Yasuda)

Abstract. A utility treatment is studied in the framework of discounted Markov decision processes. We will define a new index called a utility deviation related to the risk premium, which is characterized by an iterative formula. Examples are given in the quadratic and the exponential utility cases.

1. Introduction

This paper is concerned with the risk premium in finite Markov decision processes (MDP's) with general utility. In the utility theory, the risk premium for an arbitrary risk is defined as expected monetary value minus the amount for which a decision maker would exchange the risk, which presents a measure of aversion to the risk.

It is known by Fishburn[3] and Pratt[8] that the greater the risk aversion is, the larger the risk premium is. Thus, for the utility analysis of a stochastic process, it is meaningful to examine the risk premium associated with each policy in detail. For a utility optimization of MDP's, see our preceding paper [5] and [1, 2, 4, 7, 10, 11].

Here, in the framework of MDP's with general utility we introduce a new index, called *a utility deviation*, by which the risk premium can be characterized also. Differing from the risk premium, it is possible to approach the utility deviation by an operator, which leads us to the analysis of the iterative formula and the fixed point theory. The method employed here is closely related to the one in Sobel[10], White[11] and Chung and Sobel[1].

Section 2 will define a utility deviation on an arbitrary risk and derive its relations to the risk premium. Section 3 will prepare several notations and describe the problem concerning with a utility deviation in MDP's. Section 4 will show that an iterative formula supplied by MDP's will characterize the utility deviation.

2. Risk premium and utility deviation

In Section 2, we shall define a utility deviation for an arbitrary risk and examine its relations to the risk premium.

Consider a decision maker with a utility function g, where g is a Borel measurable function from the set of real numbers to itself. A random variable $\widetilde{\mathcal{B}}$ is called a risk, if it is non-degenerate and both $E(\mathcal{B})$ and $E(g(\mathcal{B}))$ are finite. For a risk $\mathcal{B}$, his risk premium $\sigma = \sigma(g, \mathcal{B})$ is given by

$$g\left(E(\mathcal{B}) - \sigma\right) = E\left(g(\mathcal{B})\right). \tag{2.1}$$

The equality (2.1) means that he would be indifferent between receiving the risk $\mathcal{B}$ and receiving the amount $E(\mathcal{B}) - \sigma$ (see Fishburn[3] and Pratt[8] in detail).

Now we shall define a new index $\kappa(g, \widetilde{\mathcal{B}})$ by

$$\kappa(g, \widetilde{\mathcal{B}}) = E\left(g(\widetilde{\mathcal{B}})\right) - g\left(E(\widetilde{\mathcal{B}})\right), \tag{2.2}$$

which will be called *a utility deviation.*

In the arguments of the present section, we need an assumption that *the utility function g is strictly increasing and continuous.* The assumption assures that the risk premium uniquely exists for $\mathcal{B}$. In the following Lemma 2.1 and Propositions 2.1, 2.2, we assume this assumption, however it is not spelled out.

Lemma 2.1 shows a relation between the risk premium and the utility deviation.

Lemma 2.1. *It holds that*

$$\sigma(g, \mathcal{B}) = \kappa(g^{-1}, g(\mathcal{B})). \tag{2.3}$$

Proof. Since g is strictly increasing, σ is rewritten by

$$\begin{aligned} \sigma(g, \mathcal{B}) &= E(\mathcal{B}) - g^{-1}\left(E(g(\mathcal{B}))\right) \\ &= E\left(g^{-1}g(\mathcal{B})\right) - g^{-1}\left(E(g(\mathcal{B}))\right) = \kappa(g^{-1}, g(\mathcal{B})). \end{aligned}$$

□

Propositions 2.1 and 2.2 describe the relations among the class of functions for the risk premium and the utility deviation. Pratt[8] gives several equivalent conditions to Proposition 2.2(i) with the C^2-class utility function. Proposition 2.1 easily follows from Jensen's inequality.

Proposition 2.1.

(i) *If g is (strictly) concave, $\sigma(g,\mathcal{B}) \geq (>)0$ and $\kappa(g,\mathcal{B}) \leq (<)0$ for any risk $\mathcal{B}$.*

(ii) *If g is (strictly) convex, $\sigma(g,\mathcal{B}) \leq (<)0$ and $\kappa(g,\mathcal{B}) \geq (>)0$ for any risk $\mathcal{B}$.*

(iii) *If g is linear, $\sigma(g,\mathcal{B}) = \kappa(g,\mathcal{B}) = 0$ for any risk $\mathcal{B}$.*

Let g_1 and g_2 be the utility functions. According to Nielsen[6], g_1 is called *less risk averse* than g_2 if $g_2(c) \leq E(g_2(\mathcal{B}))$ holds for a risk $\mathcal{B}$ and a real number c, then $g_1(c) \leq E(g_1(\mathcal{B}))$. Notice that $g(c) \leq E(g(\mathcal{B}))$ implies $c \leq g^{-1}(E(g(\mathcal{B})))$, so that $\sigma(g,\mathcal{B}) \leq E(\mathcal{B}) - c$ is equivalent to $g(c) \leq E(g(\mathcal{B}))$. This fact will be used in the proof of Proposition 2.2 bellow.

Proposition 2.2. *The following* (i)$\sim$(iv) *are equivalent.*

(i) $\sigma(g_1,\mathcal{B}) \leq \sigma(g_2,\mathcal{B})$ *for any risk* $\mathcal{B}$;

(ii) $\kappa(g_1^{-1}, g_1(\mathcal{B})) \leq \kappa(g_2^{-1}, g_2(\mathcal{B}))$ *for any risk* $\mathcal{B}$;

(iii) g_1 *is less risk averse than* g_2;

(iv) $g_2\, g_1^{-1}$ *is concave.*

Proof. Substitute $c = E(\tilde{\mathcal{B}}) - \sigma(g_2,\tilde{\mathcal{B}})$ to $\sigma(g_i,\tilde{\mathcal{B}}) \leq E(\tilde{\mathcal{B}}) - c$ for $i = 1,2$. Then, (iii) implies (i) from the equivalence described just before this proposition. (ii) is equivalent to (iv), since they are equivalent to $E(g_2 g_1^{-1}(\tilde{\mathcal{B}})) \leq g_2 g_1^{-1}\left(E(\tilde{\mathcal{B}})\right)$ for any $\tilde{\mathcal{B}}$. The other proofs follow easily from (2.1), (2.2) and (2.3). □

3. Description of the problem

The previous section has shown the validity of the utility deviation on the risk $\mathcal{B}$. In this section, we shall define a utility deviation on MDP's with the general utility.

We consider the standard MDP's specified by (S, A, P, r, β), where $S = \{1, 2, \cdots, N\}$ is a finite state space, A is an action space, $P = (p_{ij}^a)$ is the matrix of transition probabilities satisfying that $p_{ij}^a \geq 0$, $\sum_{j\in S} p_{ij}^a = 1$ for all $i \in S, a \in A$, $r(i,a)$ is an immediate reward function defined on $S \times A$ and $\beta (0 < \beta < 1)$ is a discount factor. Assume that A is a Borel set, r is bounded measurable and $r(i,a) \geq 0$ for all $i \in S, a \in A$.

The sample space is the product space $\Omega = (S \times A)^\infty$ such that the projections X_t, Δ_t to the t-th factors S, A describe the state and the action of the process at time

$t \geq 0$, respectively. We treat only the randomized stationary policy, which is defined by a conditional probability $\pi(\cdot|i)$ on A for each $i \in S$. The set of all randomized stationary policies is denoted by Π. Let $H_t = (X_0, \Delta_0, \cdots, \Delta_{t-1}, X_t)$ for $t \geq 0$. We assume that, for each $\pi \in \Pi$ with $t \geq 0, i, j \in S$ and $a \in A$,

$$Prob(\Delta_t = a | H_{t-1}, \Delta_{t-1}, X_t = i) = \pi(a|i),$$
$$Prob(X_{t+1} = j | H_{t-1}, \Delta_{t-1}, X_t = i, \Delta_t = a) = p_{ij}^a.$$

Then, the initial state $i \in S$ and the policy $\pi \in \Pi$ determine the probability measure P_i^π on Ω by a usual way.

The present value of the state-action process $(X, \Delta) = \{(X_t, \Delta_t);\ t = 0, 1, 2, \cdots\}$ is defined by

$$\mathcal{B}_{X,\Delta} := \sum_{t=0}^{\infty} \beta^t r(X_t, \Delta_t).$$

Let g be a utility function bounded bellow, evaluating the present value. Since $g(x)$ is equivalent to $ag(x) + b$ for any constants $a > 0$ and b, we may assume without loss of generality that g is a function from the interval $[0, \infty)$ to itself.

We define *the utility deviation* κ_i^π of g for any initial state i and policy $\pi \in \Pi$ by

$$\kappa_i^\pi := E_i^\pi(g(\mathcal{B}_{X,\Delta})) - g\left(E_i^\pi(\mathcal{B}_{X,\Delta})\right), \tag{3.1}$$

where E_i^π is the expectation with respect to P_i^π. Let the distribution functions of $\mathcal{B}_{X,\Delta}$ for $i \in S$ be

$$F_i^\pi(x) := P_i^\pi(\mathcal{B}_{X,\Delta} \leq x) \quad \text{for } x \in [0, \infty). \tag{3.2}$$

then, (3.1) is written by

$$\kappa_i^\pi = \int_0^\infty g(x)\, dF_i^\pi(x) - g\left(\int_0^\infty x\, dF_i^\pi(x)\right). \tag{3.3}$$

Our problem is to give a characterization for the utility deviation κ_i^π, which will be investigated in the next section.

4. Characterization of utility deviation on MDP's

In this section, the utility deviation will be characterized by an iterative formula.

The utility deviation κ_i^π is given by (3.1) or (3.3) for each policy $\pi \in \Pi$ and the initial state i associated with MDP (S, A, P, r, β) in the previous section. Suppressing

this fixed π for the sake of brevity, we shall give several notations. For $i \in S$, let

$$r_i := \sum_{a \in A} r(i,a)\pi(a|i),$$

$$q_{ij} := \sum_{a \in A} p_{ij}^a \pi(a|i) \quad \text{and}$$

$$\varphi_i := \int_0^\infty x \, dF_i(x) \quad \text{where } F_i(x) = F_i^\pi(x).$$

Note that φ_i represents the expected total discounted reward in case of a linear utility function. Therefore the following is well known results in the theory of Markov decision processes.

Lemma 4.1. (Ross[9]) *The expected total discounted reward* $\{\varphi_i : i \in S\}$ *is the unique bounded solution of the equation:*

$$\varphi_i = r_i + \beta \sum_{j \in S} q_{ij}\varphi_j \quad \textit{for} \quad i \in S.$$

The following result is given by Sobel[10] and will be used in the proof of Theorem 4.1 bellow.

Lemma 4.2. (Sobel[10]) *For any* $\pi \in \Pi$ *and* $i \in S$, *it holds that*

$$F_i(x) = \sum_{j \in S} q_{ij} F_j\left((x - r_i)/\beta\right). \tag{4.1}$$

In order to characterize the utility deviation κ_i^π by an operator on some family of probability distributions, we prepare the following notations. Let

$$\Psi := \{\, G \,|\, G \text{ is the probability distribution on } [0,\, M_\beta] \,\},$$

where $M_\beta := M/(1-\beta)$ and $M := \sup_{\{i \in S, a \in A\}} r(i,a)$. Let $\mathcal{L} := \times_{i \in S} \Psi$ be the product space. Associated with each $i \in S$ is an operator $T^i : \mathcal{L} \longrightarrow \mathcal{L}$ defined as follows : For $G(x) = (G_j(x) \; ; \; j \in S) \in \mathcal{L}$, let

$$\begin{aligned} T^i(G) &= (T^i(G)_j \; ; \; j \in S) \quad \text{and} \\ T^i(G)_j(x) &= G_j\left((x - r_i)/\beta\right). \end{aligned} \tag{4.2}$$

Since $0 \le r_i \le M$ and $G_j\left((M_\beta - r_i)/\beta\right) \ge G_j((M_\beta - M)\ /\beta) = G_j(M_\beta) = 1$, it follows $(T^i(G)_j; j \in S) \in \mathcal{L}$. Then, the operator T^i is well defined. Notice from (4.2) that $T^{i_1}T^{i_2}\cdots T^{i_n}(G)_j(x)$ means $T^{i_1}(T^{i_2}\cdots T^{i_n}(G))_j(x)$ where $T^{i_2}\cdots T^{i_n}(G) \in \mathcal{L}$.

We extend the domain of κ_i^π to $\mathcal{L}$ component-wise: For $G = (G_i;\, i \in S) \in \mathcal{L}$, let

$$\begin{aligned} &\kappa(G) := (\,\kappa(G_i)\,;\, i \in S\,) \quad \text{and} \\ &\kappa(G_i) := \int_0^{M_\beta} g(x)\, dG_i(x) - g\left(\int_0^{M_\beta} x\, dG_i(x)\right). \end{aligned} \tag{4.3}$$

The policy π determins $F = (F_i;\, i \in S) \in \mathcal{L}$ and $\kappa(F) = (\kappa(F_i)\,;\, i \in S\,)$, where $F_i(x) = F_i^\pi(x)$ is given by (3.2). It is clear from (3.3) and (4.3) that $\kappa(F_i) = \kappa_i^\pi$ for $i \in S$.

Now, the utility deviation $\kappa(F)$ is presented by an iterative formula in the following theorem.

Theorem 4.1. *For any fixed $\pi \in \Pi$, $\kappa(F) = (\kappa(F_i)\,;\, i \in S\,)$ satisfies the following equations:*

$$\kappa(F_i) = \bar{g}_i + \textstyle\sum_{j\in S} q_{ij}\kappa\,(T^i(F)_j) \quad \textit{and} \tag{4.4}$$

$$\kappa\left(T^{i_1}T^{i_2}\cdots T^{i_n}(F)_i\right) = \bar{g}_{i_1,i_2,\cdots i_n,i} + \sum_{j\in S} q_{ij}\,\kappa\left(T^{i_1}T^{i_2}\cdots T^{i_n}T^i(F)_j\right) \tag{4.5}$$

for any i, i_1, i_2, $\cdots$, $i_n \in S$, $n \geq 1$, where

$$\bar{g}_i = \textstyle\sum_{j\in S} q_{ij} g\,(r_i + \beta\varphi_j) - g(\varphi_i) \quad \textit{and}$$

$$\begin{aligned} \bar{g}_{i_1,i_2,\cdots i_n,i} = \textstyle\sum_{j\in S} q_{ij} g\,(&r_{i_1} + \beta r_{i_2} + \cdots + \beta^{n-1} r_{i_n} + \beta^n r_i + \beta^{n+1}\varphi_j) \\ &- g\,(r_{i_1} + \beta r_{i_2} + \cdots + \beta^{n-1} r_{i_n} + \beta^n \varphi_i)\,. \end{aligned}$$

Proof. We prove (4.5) in case of $n = 1$. The other cases are proved analogously. By (4.2) and (4.3), we have

$$\kappa\left(T^{i_1}(F)_i\right) = \int_0^{M_\beta} g(x)\, dT^{i_1}(F)_i(x) - g\left(\int_0^{M_\beta} x\, dT^{i_1}(F)_i(x)\right). \tag{4.6}$$

Since $T^{i_1}(F)_i(x) = F_i\,((x - r_{i_1})/\beta)$, it holds from (4.1) that

$$T^{i_1}(F)_i(x) = \sum_{j\in S} q_{ij} F_j\left((x - r_{i_1} - \beta r_i)/\beta^2\right) = \sum_{j\in S} q_{ij}\left(T^{i_1}T^i(F)_j\right)(x).$$

Therefore,

$$\begin{aligned} &\int_0^{M_\beta} g(x)\, dT^{i_1}(F)_i(x) = \sum_{j\in S} q_{ij} \int_0^{M_\beta} g(x)\, d\left(T^{i_1}T^i(F)_j\right)(x), \\ &\int_0^{M_\beta} x\, d\left(T^{i_1}T^i(F)_j\right)(x) = r_{i_1} + \beta r_i + \beta^2\varphi_j \quad \text{and} \\ &\int_0^{M_\beta} x\, dT^{i_1}(F)_i(x) = r_{i_1} + \beta\varphi_i\,. \end{aligned}$$

Using these facts, we get from (4.6) that

$$\begin{aligned}\kappa\left(T^{i_1}(F)_i\right) &= \sum_{j\in S} q_{ij}\left[\int_0^{M_\beta} g(x)\, d\left(T^{i_1}T^i(F)_j\right)(x)\right. \\ &\qquad \left. -g\left(\int_0^{M_\beta} x\, d\left(T^{i_1}T^i(F)_j\right)(x)\right)\right] \\ &\quad +\sum_{j\in S} q_{ij} g\left(\int_0^{M_\beta} x\, d\left(T^{i_1}T^i(F)_j\right)(x)\right) \\ &\qquad -g\left(\int_0^{M_\beta} x\, dT^{i_1}(F)_i(x)\right) \\ &= \sum_{j\in S} q_{ij}\kappa\left(T^{i_1}T^i(F)_j\right) + \bar{g}_{i_1,i},\end{aligned}$$

which implies (4.5) in case of $n = 1$. □

The evaluation of $\kappa(F_i)$ could be obtained iteratively using Theorem 4.1.

Corollary 4.1. *Let* $\kappa_i^{(1)} := \bar{g}_i$ *and, for* $n \geq 1$,

$$\kappa_i^{(n+1)} := \kappa_i^{(n)} + \sum_{i_1,i_2,\cdots,i_n\in S} q_{i,i_1}q_{i_1,i_2}\cdots q_{i_{n-1},i_n}\bar{g}_{i,i_1,i_2,\cdots,i_n}.$$

If the utility $g(x)$ *is differentiable and* $|g'(x)| \leq L$ *on* $[0,\ M_\beta]$ *for some* $L > 0$, *then we get from* (4.4) *and* (4.5),

$$|\kappa(F_i) - \kappa_i^{(n)}| \leq \beta^n M_\beta L \quad \textit{for each } n \geq 1.$$

We shall give examples illustrating Theorem 4.1.

Example 1. Consider the case of $g(x) = x^2$. Since $\kappa(F_i) = \kappa_i^\pi = E_i^\pi\left(\mathcal{B}_{X,\Delta}^2\right) - \left(E_i^\pi(\mathcal{B}_{X,\Delta})\right)^2$, the utility deviation is equal to the variance of the present value. In this case, we get

$$\kappa\left(T^i(F)_j\right) = \beta^2\kappa\left(F_j\right) \quad \text{and}$$
$$\bar{g}_i = \textstyle\sum_{j\in S} q_{ij}(r_i + \beta\varphi_j)^2 - \varphi_i^2.$$

So, (4.4) becomes

$$\kappa(F_i) = \bar{g}_i + \beta^2\sum_{j\in S} q_{ij}\kappa(F_j) \quad \text{for} \quad i \in S. \tag{4.7}$$

Denoting $\bar{g} = (\bar{g}_i : i \in S)$, (4.7) is represented in the matrix form by

$$\kappa(F)^t = \bar{g}^t + \beta^2 Q\ \kappa(F)^t,$$

where t means a transpose of the vector. Therefore,

$$\kappa(F)^t = [I - \beta^2 Q]^{-1} \bar{g}^t,$$

which is the same expression as that obtained in Theorem 1 of Sobel[10].

Example 2. Consider the exponential utility case, i.e., $g_\lambda(x) = 1 - \exp(-\lambda x)$ for $\lambda > 0$. The utility deviation will be denoted by $\kappa(\lambda, F_i) = \kappa_i^\pi$ with $\kappa(g_\lambda, \mathcal{B}_{X,\Delta}) = (\kappa_i^\pi; i \in S)$. After some simple calculations, we get

$$\kappa\left(\lambda,\, T^i(F)_j\right) = e^{-\lambda r_i} \kappa(\beta\lambda,\, F_j)$$

and

$$\bar{g}_i = e^{-\lambda \varphi_i} - \sum_{j \in S} \bar{q}_{ij} e^{-\beta\lambda\varphi_j}$$

where $\bar{q}_{ij} = q_{ij} e^{-\lambda r_i}$. So, (4.4) becomes

$$\kappa\,(\lambda,\, F_i) = \bar{g}_i + \sum_{j \in S} \bar{q}_{ij} \kappa(\beta\lambda,\, F_j) \tag{4.8}$$

for each $i \in S$. Using (4.8), we can find the method of successive approximation for obtaining $\kappa\,(\lambda,\, F)$.

References

[1] K.J. Chung and M.J. Sobel, Discounted MDP's: Distribution functions and exponential utility maximization, *SIAM J. Control and Optimization*, **25**(1987), 49-62.

[2] E.V. Denardo and U.G. Rothblum, Optimal stopping, exponential utility and linear programming, *Math. Prog.*, **16** (1979), 228–244.

[3] P.C. Fishburn, Utility Theory for Decision Making, John Wiley & Sons, New York, 1970.

[4] R.S. Howard and J.E. Matheson, Risk-sensitive Markov decision processes, *Manag. Sci.*, **8** (1972), 356–369.

[5] Y. Kadota, M. Kurano and M. Yasuda, Discounted Markov decision processes with general utility functions, *Proc. of APORS'94*, World Scientific, 1995, 330–337.

[6] L.T. Nielsen, The expected utility of portfolios of assets, *J. Math. Economics*, **22** (1993), 439–461.

[7] E.L. Porteus, On the optimality of structured policies in countable stage decision processes, *Manag. Sci.*, **22** (1975), 148–157.

[8] J.W. Pratt, Risk aversion in the small and in the large, *Econometrica*, **32** (1964) 122–136.

[9] S.M. Ross, Applied Probability Models with Optimization Applications, Holden-Day, 1970.

[10] M.J. Sobel, The variance of discounted Markov decision processes, *J. Appl. Prob.*, **19** (1982) 794–802.

[11] D.J. White, Minimizing a threshold probability in discounted Markov decision processes, *J. Math. Anal. Appl.*, **173** (1993) 634–646.