Fuzzy perceptive values for stopping models and MDPs

01101550 千葉大学教育学部 蔵野正美 KURANO Masami
01701690 千葉大学理学部 安田正實 YASUDA Masami
01401530 千葉大学理学部 *中神潤一 NAKAGAMI Jun-ichi
01702986 北九州市立大学経済学部 吉田祐治 YOSHIDA Yuji

1. Introduction

In a real application of such a mathematical model as a optimal stopping or a Markov decision process (MDP), it often occurs that the required data is linguistically and roughly perceived (for example, the price of the asset is about \$100, etc). A possible way of handling such a perception-based information is to use the fuzzy set (cf. [3]), whose membership function describes the perception value of the required data. If the fuzzy perception of the required data is given, how can we estimate the future expected reward, called a fuzzy perceptive value, under the condition that we can know the true value of the required data immediately before our decision making. The problem formulation is inspired by Zadeh's paper [6], in which the perception-based-theory of probabilistic reasoning is developed and the idea of the perceptive value(possibility distribution) of the objective function under the possibility constraints is proposed by using a generalized extension principle.

Here, we formulate the perceptive models for optimal stopping problems and MDPs, and the corresponding fuzzy perceptive values are characterized and calculated by a fuzzy optimality equations. Numerical examples are given.

Let $\mathbb{R}$ be the set of all real numbers and $\widetilde{\mathbb{R}}$ the set of all fuzzy numbers, i.e., $\widetilde{s} \in \widetilde{\mathbb{R}}$ means that $\widetilde{s} : \widetilde{\mathbb{R}} \to [0,1]$ is normal, upper-semicontinuous and fuzzy convex and has a compact support, The α-cut of $\widetilde{s} \in \widetilde{\mathbb{R}}$ is given by $\widetilde{s}_\alpha := \{x \in \mathbb{R} \mid \widetilde{s} \geq \alpha\}(\alpha \in (0,1])$ and $\widetilde{s}_0 := \mathrm{cl}\{x \in \mathbb{R} \mid \widetilde{s} > 0\}$, where $\mathrm{cl}A$ is the closure of a set A. We write $\widetilde{s} = [\widetilde{s}_\alpha^-, \widetilde{s}_\alpha^+]\alpha \in [0,1]$. For $\widetilde{s}, \widetilde{r} \in \widetilde{\mathbb{R}}$,

$$\widetilde{\max}\{\widetilde{s}, \widetilde{r}\}(y) := \sup_{\substack{x_1, x_2 \in \mathbb{R} \\ y = x_1 \vee x_2}} \{\widetilde{s}(x_1) \wedge \widetilde{r}(x_2)\} \quad (y \in \mathbb{R}),$$

then $\widetilde{s} \preccurlyeq \widetilde{r}$ (fuzzy max order) means $\widetilde{r} = \widetilde{\max}\{\widetilde{s}, \widetilde{r}\}$.

2. Perceptive stopping model

Let $\mathfrak{X}$ be the set of all integrable random variables on the probability space $(\Omega, \mathcal{M}, P)$ and $\mathfrak{X}^n = \{X = (X_1, \ldots, X_n) \mid X_t \in \mathfrak{X}\ (1 \leq t \leq n)\}$. For each sequence of random variables $X = (X_1, \ldots, X_n) \in \mathfrak{X}^n$, we denote by $\delta^* = \delta^*(X)$ the optimal stopping time for X (cf. [1]) with the optimal expected reward $E(X_{\delta^*}) := E(X_{\delta^*(X)})$.

A measurable map $\widetilde{X} : \Omega \to \widetilde{\mathbb{R}}$ is called a fuzzy perception on $\mathfrak{X}$. For a sequence of fuzzy perceptions $\widetilde{X} = \widetilde{X}_1, \ldots, \widetilde{X}_n$, the problem is to characterize and compute the perceptive value $E\widetilde{X}_{\delta^*}$ where

$$E\widetilde{X}_{\delta^*}(x) = \sup_{\substack{x = E(X_{\delta^*}) \\ X \in \mathfrak{X}^n}} \widetilde{X}(X), \tag{2.1}$$

$$\widetilde{X}(X) = \sup_{\omega \in \Omega} \widetilde{X}_1(\omega)(X_1(\omega)) \wedge \cdots \wedge \widetilde{X}_n(\omega)(X_n(\omega)) \quad (a \wedge b = \min\{a, b\}). \tag{2.2}$$

Theorem 2.1 *The following holds:*

(i) $E(\widetilde{X}_\delta) \in \widetilde{\mathbb{R}}$.

(ii) Suppose that $\widetilde{X} = (\widetilde{X}_1, \ldots, \widetilde{X}_n)$ *is independent with each* $\widetilde{X}_t$ $(t = 1, 2, \ldots, n)$. *Then* $E\widetilde{X}$ *is given by the backward recursive equation.*

$$\widetilde{\gamma}_n^n = E(\widetilde{X}_n), \quad \widetilde{\gamma}_k^n = E\widetilde{\max}\{\widetilde{X}_k, \widetilde{\gamma}_{k+1}^n \ (k = n-1, \ldots, 2, 1) \text{ and } \widetilde{\gamma}_1^n = E(\widetilde{X}_{\delta^*}). \tag{2.3}$$

3. Perceptive MDPs

Consider finite state and action spaces, S and A, containing $n < \infty$ and $k < \infty$ elements with $S = \{1, 2, \ldots, n\}$ and $A = \{1, 2, \ldots, k\}$. Let $\mathcal{P}(S) \subset \mathbb{R}^n$ and $\mathcal{P}(S|SA) \subset \mathbb{R}^{n \times nk}$ be the sets of all probabilities on S and conditional probabilities on S given $S \times A$, that is,

$$\begin{aligned} \mathcal{P}(S) &:= \{q = (q(1), q(2), \ldots, q(n))' \mid q(i) \geq 0, \sum_{i=1}^{n} q(i) = 1,\ i \in S\}, \\ \mathcal{P}(S|SA) &:= \{Q = (q_{ia}(\cdot) : i \in S, a \in A) \mid \\ & \qquad q_{ia}(\cdot) = (q_{ia}(1), q_{ia}(2), \ldots, q_{ia}(n))' \in \mathcal{P}(S),\ i \in S, a \in A\}. \end{aligned}$$

For any $Q = (q_{ia}(\cdot)) \in \mathcal{P}(S|SA)$, A MDP is specified by $\{S, A, Q, r\}$, where $r : S \times A \to \mathbb{R}_+$ is an immediate reward function. Denote by F the set of functions from S to A. A policy π is a sequence $(f_1, f_2, \ldots)$ of functions with $f_t \in F$ $(t \geq 1)$. Let Π denote the class of policies. We denote by f the policy $(f_1, f_2, \ldots)$ with $f_t = f$ for all $t \geq 1$ and some $f \in F$. Such a policy is called stationary.

We associate with each $f \in F$, $Q \in \mathcal{P}(S|SA)$ the column vector $r(f) = (r(1, f(1)), \ldots, r(n, f(n)))'$ and the $n \times n$ transition matrix $Q(f)$, whose (i, j) element is $q_{i,f(i)}(j)$ $1 \leq i, j \leq n$. Then, the expected total discounted reward from $\pi = (f_1, f_2, \ldots)$ is the column vector $\psi(\pi|Q) = (\psi(1, \pi|Q), \ldots, \psi(n, \pi|Q))'$, which is defined, as a function of $Q \in \mathcal{P}(S|SA)$, by

$$\psi(\pi|Q) = \sum_{t=0}^{\infty} \beta^t Q(f_1) Q(f_2) \cdots Q(f_t) r(f_{t+1}), \tag{3.1}$$

where $0 < \beta < 1$ is a discount factor.

It is well-known (cf. [4]) that for each $Q \in \mathcal{P}(S|SA)$, a optimal stationary policy $f^* = f^*(Q)$ exists with

$$\psi(i, f^*|Q) = \sup_{\pi \in \Pi} f(i, \pi|Q) := \psi^*(i, Q). \tag{3.2}$$

The perception $\widetilde{Q}_{ia}$ on $\mathcal{P}(S)$ is supposed to be given for each $i \in S, a \in A$. Then the perception $\widetilde{Q}$ on $\mathcal{P}(S|SA)$ is defined by

$$\widetilde{Q}(Q) = \min_{i \in S,\ a \in A} \widetilde{Q}_{ia}(q_{ia})(\cdot), \text{ where } Q = (q_{ia} : i \in S, a \in A) \in \mathcal{P}(S|SA). \tag{3.3}$$

The problem is to find the perception value $\widetilde{\psi}(i)$, where

$$\widetilde{\psi}(i)(x) = \sup_{Q \in \mathcal{P}(S|SA),\ x = \psi^*(i,Q)} \widetilde{Q}(Q). \tag{3.4}$$

Theorem 3.1 *The following holds:*

(i) $\widetilde{\psi}(i) \in \widetilde{\mathbb{R}}$.

(ii) $\widetilde{\psi}(i)$ $(i \in S)$ *is given as a unique solution of the fuzzy optimality equation:*

$$\widetilde{\psi}(i) = \widetilde{\max}\{\mathbf{1}_{\{r(i,a)\}} + \beta \widetilde{Q}_{ia} \cdot \widetilde{\psi}\}, \tag{3.5}$$

where $\widetilde{Q} \cdot \widetilde{\psi}(x) = \sup \widetilde{Q}_{ia}(q) \wedge \widetilde{\psi}(\psi)$ *and the supremum is taken on the range* $\{(q, \psi) \in \mathcal{P}(S) \times \mathbb{R}^n \mid x = \sum_{j=1}^{n} q(j)\psi_j\}$ *and* $\widetilde{\psi}(\psi) = \widetilde{\psi}(1)(\psi_1) \wedge \cdots \wedge \widetilde{\psi}(n)(\psi_n)$ *with* $\psi = (\psi_1, \ldots, \psi_n) \in \mathbb{R}^n$.

References

[1] Chow, Y. S., Robbins, H. and Siegmund, D., *The Theory of Optimal Stopping : Great Expectations.* (1971), New York: Houghton Mifflin.

[2] Kurano, M., Yasuda, M., Nakagami, J. and Yoshida, Y., A fuzzy stopping problem with the concept of perception. *Fuzzy Optimization and Decision Making*, (2004), **3**, 367–374.

[3] Kurano, M., Yasuda, M., Nakagami, J. and Yoshida, Y., Fuzzy perceptive values for MDPs with discounting. (preprint).

[4] Mine, H. and Osaki, S., *Markov Decision Processes.* (1970), Elsevier, Amsterdam.

[5] Zadeh, L. A., Fuzzy sets. *Inform. and Control*, (1965), **8**, 338–353.

[6] Zadeh, L. A., Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. *J. of Statistical Planning and Inference*, (2002), **105**, 233–264.