Journal of Uncertain Systems
Vol.7, No.1, pp.50-57, 2013
Online at: www.jus.org.uk

Conditional Value-at-Risk Minimization in Finite State Markov Decision Processes: Continuity and Compactness

Buheeerdun Yang*
Graduate School of Science, Chiba University, Japan

Received 28 September 2011; Revised 2 June 2012

Abstract

This study is concerned with the dynamic risk-analysis for finite state Markov decision processes. As a measure of risk, we consider conditional value-at-risk(CVaR) for the real value of the discounted total reward from a policy, under whose criterion risk optimal or deterministic policies are defined. The risk problem is equivalently redefined as a non-linear optimization problem on the attainable set of the distribution functions for the real values over all policies. Showing the weak-continuity of CVaR on the space of attainable distribution functions, the mathematical existence theorem of optimal policies are proved throughout the discussion of convex analysis and weak-compactness.

Keywords: risk-optimization, Markov decision processes, conditional value-at-risk, discounted case, finite state and action, convex analysis, existence theorem

1 Introduction and Notation

In probabilistic methods in financial engineering [8], many risk measures have been generalized and analyzed in economically motivated optimization problems, for example, value-at-risk (VaR), conditional value-at-risk (CVaR) [19, 20], coherent risk of measure [1, 11, 12], convex risk of measure [7, 6] and its applications [9, 21]. In particular, CVaR has good properties and easy to analyze such complex models as sequential decision processes because it can be expressed by a remarkable minimization formula. As for sequential decision processes, Markov decision processes (MDPs) have many applications in such wide fields as ecology, economics and communications engineering [16]. So it is important to analyze the dynamic risk-model for MDPs, whose studies are done by many authors [14, 23, 22], in which the real value of total reward from a policy is evaluated by the target-percentile risk measure. The key feature of analysis is to characterize the risk-optimal policy and its corresponding value functions by dynamic programming methods. Applying the results of these studies, Boda and Filar [3] has considered the risk problem for MDPs using "Value-at-Risk (VaR)" or "Conditional Value-at-Risk (CVaR)" as a measure of risk. Also, Mundt [13] has solved an optimization problem for a risk-management model (dynamic portfolio optimization) applying the theory of MDPs. In this paper, we provide an alternative framework for the study of risk minimization problem in finite state MDPs, which is solved based on convex analysis including continuity and compactness [4]. We consider CVaR for the real value of the discounted total reward from a policy, which is minimized over all policies. Observing that CVaR of a random income variable is depending only on its distribution, the problem to examined will be equivalently reformulated as non-linear optimization one on the attainable set of distribution functions for the real values over all policies, which enables the discussion of convex analysis and compactness. Showing the continuity of CVaR with respect to the distribution function, we will give the existence theorem for risk-optimal or deterministic risk-optimal policies. In the remainder, we shall establish notation that is used throughout the paper and define the problem to be considered.

Let $\mathbb{R}$ be the set of all real numbers. Let X be a random reward variable on some probability space $(\Omega, \mathcal{B}, P)$, and $F_X(x)$ the distribution function, i.e., $F_X(x) = P(X \leq x)$ $(x \in \mathbb{R})$. The inverse function $F_X^{-1}(p)$ $(0 < p < 1)$ will be defined by $F_X^{-1}(p) = \inf\{x \in \mathbb{R} | F_X(x) \geq p\}$. Then, the conditional value-at-risk

*Corresponding author. Email: bheed20020330@graduate.chiba-u.jp (B. Yang).

for a level $\gamma \in (0,1)$ of X, $\mathrm{CVaR}_\gamma(X)$, is defined [19, 20] by

$$\mathrm{CVaR}_\gamma(X) = \frac{1}{1-\gamma}\int_\gamma^1 F_{-X}^{-1}(p)dp. \tag{1}$$

We need the following two propositions used in the sequel.

Proposition 1.1 (Fundamental minimization formula) [20] *For any random variable X with $E|X| < \infty$, $\mathrm{CVaR}_\gamma(X)$ is finite for $\gamma \in (0,1)$ and it holds that*

$$\mathrm{CVaR}_\gamma(X) = \inf_{b\in\mathbb{R}}\left\{b + \frac{1}{1-\gamma}E[(-X-b)^+]\right\}. \tag{2}$$

Moreover, the infimum is attained at $b^* = F_{-X}^{-1}(\gamma)$.

For any Borel subset D of $\mathbb{R}$, $P(D)$ denotes the set of all probability measures on D.

Proposition 1.2 [2, 10, 17] *Let $\mathcal{G}$ be a family of real-valued measurable function on a Borel set D of $\mathbb{R}$. Then, if $\mathcal{G}$ is uniformly bounded and equicontinuous at each $x \in D$, for any sequence $\{P_n\}$ in $P(D)$ which converges weakly to $P \in P(D)$, it holds that*

$$\sup_{g\in\mathcal{G}}\left|\int g dP_n - \int g dP\right| \to 0 \ \ as \ \ n \to \infty.$$

We consider standard MDPs specified by (S, A, q, r), where both S and A are finite sets and denote the sets of the state of the process and actions available at each state, respectively, and $q = (q_{ij}(a))$ is the matrix of transition probabilities satisfying that $\sum_{j\in S} q_{ij}(a) = 1$ for all $i \in S$ and $a \in A$, $r(i,a)$ is an immediate reward function defined on $S \times A$. The sample space is the product space $\Omega = (S \times A)^\infty$ such that the projection X_t, Δ_t on the $t-$th factor S, A describe the state and the action of $t-$time of the process $(t \geq 0)$. Let $\Omega_t = (S \times A)^t \times S$ $(t \geq 0)$. A policy $\pi = (\pi_0, \pi_1, \dots)$ is a sequence of conditional probability π_t such that $\pi_t(A|h_t) = 1$ for all $t \geq 0$ and $h_t = (x_0, a_0, x_1, a_1, \dots, x_t) \in \Omega_t$. The set of all policies is denoted by Π. Let $H_t = (X_0, \Delta_0, \dots, \Delta_{t-1}, X_t) \in \Omega_t$ for $t \geq 0$.

We assume that for each $\pi = (\pi_0, \pi_1, \dots) \in \Pi, P_\pi(X_{t+1} = j|H_{t-1}, \Delta_{t-1}, X_t = i, \Delta_t = a) = q_{ij}(a)$ for all $t \geq 0$, $i, j \in S$, $a \in A$. Then, for any initial probability measure $\nu \in P(S)$ and policy $\pi \in \Pi$, we can define the probability measure P_π^ν on Ω in an obvious way. The expectation with respect to P_π^ν is denoted by E_π^ν. Also, the initial distribution degenerate at state $i \in S$ is denoted simply by i. The real value of the discounted total reward of the state-action processes $\{X_t, \Delta_t : t = 0, 1, 2, \dots\}$ under a policy $\pi \in \Pi$ is defined by

$$\tilde{\varphi}_\pi^\nu = \lim_{T\to\infty} \tilde{\varphi}_\pi^{\nu,T}, \tag{3}$$

where

$$\tilde{\varphi}_\pi^{\nu,T} = \sum_{t=0}^{T} \beta^t r(X_t, \Delta_t). \tag{4}$$

The problem is to minimize the conditional value-at-risk of $\tilde{\varphi}_\pi, \mathrm{CVaR}_\gamma(\tilde{\varphi}_\pi)$, over all policies $\pi \in \Pi$. The policy $\pi^* \in \Pi$ is said to be risk-optimal for $\nu \in P(S)$ if it holds that

$$\mathrm{CVaR}_\gamma(\tilde{\varphi}_{\pi^*}^\nu) \leq \mathrm{CVaR}_\gamma(\tilde{\varphi}_\pi^\nu) \ \ for\ any \ \ \pi \in \Pi. \tag{5}$$

For notational convenience, ν will be suppressed for P_π^ν, $\tilde{\varphi}_\pi^\nu$, $\tilde{\varphi}_\pi^{\nu,T}$ and E_π^ν here onwards.

In Section 2 by the discussion of continuity and compactness, the existence of a risk-optimal policy is shown. In Section 3, a deterministic policy is considered through convex analysis.

2 Continuity and Compactness

In this section, we give the existence theorem of risk-optimal policies. For a sequence of policies $\{\pi^n : n = 1, 2, \ldots\} \subset \Pi$ with $\pi^n = (\pi_0^n, \pi_1^n, \pi_2^n, \ldots)$ and $\pi = (\pi_0, \pi_1, \pi_2, \ldots) \in \Pi$, we say that a sequence $\{\pi^n\}$ converges to π if $\pi_t^n(a|h_t) \to \pi_t(a|h_t)$ as $n \to \infty$ for any $a \in A$, $h_t \in \Omega_t$ and $t \geq 0$.

The following facts have been proved by Derman [5].

Lemma 2.1 [5] *It holds that*

(i) *The class* Π *is compact,*

(ii) $P_\pi^\nu(H_t = h_t, \Delta_t = a_t)$ *is continuous in* $\pi \in \Pi$ *for each* $h_t \in \Omega_t$, $a_t \in A$ *and* $t \geq 0$.

Let, for each $\nu \in P(S)$ and $\pi \in \Pi$,

$$F_\pi^\nu(x) := P_\pi^\nu(-\tilde{\varphi}_\pi \leq x),$$
$$\Phi(\nu) := \{F_\pi^\nu(\cdot)|\pi \in \Pi\}.$$

Since both S and A as finite sets, there exists a constant M such that $|r(i, a)| \leq M$ for all $i \in S$ and $a \in A$. Then, $F \in \Phi(\nu)$ is a distribution function on $[\underline{M}, \overline{M}]$, where $\underline{M} := -M/(1-\beta)$ and $\overline{M} := M/(1-\beta)$.

We denote by $C[\underline{M}, \overline{M}]$ the set of all bounded and continuous functions on $[\underline{M}, \overline{M}]$.

Lemma 2.2 *For any* $g \in C[\underline{M}, \overline{M}]$, $E_\pi g(\tilde{\varphi}_\pi)$ *is continuous in* $\pi \in \Pi$.

Proof: By Lemma 2.1 (ii), $E_\pi^\nu g(\tilde{\varphi}_\pi)$ is continuous in $\pi \in \Pi$. Since g is uniformly continuous on $[\underline{M}, \overline{M}]$, for any $\varepsilon > 0$ there exist $\delta > 0$ such that $|g(x)-g(y)| < \varepsilon$ if $|x-y| < \delta$. On the other hand, $|\tilde{\varphi}_\pi^T - \tilde{\varphi}_\pi| \leq \beta^{T+1}M/(1-\beta)$, so that there exist $T^* > 0$ with $|\tilde{\varphi}_\pi^T - \tilde{\varphi}_\pi| \leq \delta$ for all $T \geq T^*$. Thus, it yields that

$$|E_\pi g(\tilde{\varphi}_\pi^T) - E_\pi g(\tilde{\varphi}_\pi)| \leq E_\pi \left(|g(\tilde{\varphi}_\pi^T) - g(\tilde{\varphi}_\pi)|\right) < \varepsilon \;\; for \;\; T \geq T^*,$$

which $E_\pi g(\tilde{\varphi}_\pi^T)$ uniformly converges to $E_\pi g(\tilde{\varphi}_\pi)$. This shows the continuity of $E_\pi^\nu g(\tilde{\varphi}_\pi)$. □

Lemma 2.3 *It holds that*

(i) *For any* $\nu \in P(S)$, $\Phi(\nu)$ *is weak-compact;*

(ii) *If a sequence* $\{\pi^n\} \subset \Pi$ *converges to* $\pi \in \Pi$ *as* $n \to \infty$, *the corresponding sequence* $\{F_{\pi^n}\} \subset \Phi(\nu)$ *weakly converges to* $F_\pi \in \Phi(\nu)$.

Proof: For any sequence $\{F_{\pi^n} : n = 1, 2, \ldots\} \subset \Phi(\nu)$, we prove that there exists a subsequence of $\{F_{\pi^n}\}$ which weakly converges to some $F \in \Phi(\nu)$. In fact, from Lemma 2.1(i), Π is compact, so that without loss of generality we can assume that there exists a $\pi \in \Pi$ such that $\{\pi^n\}$ converges to π. For any $g \in C[\underline{M}, \overline{M}]$,

$$\int g(x)dF_{\pi^n} = E_{\pi^n} g(\tilde{\varphi}_{\pi^n}) \;\; (n \geq 1)$$

and

$$\int g(x)dF_\pi = E_\pi g(\tilde{\varphi}_\pi).$$

Therefore, by Lemma 2.2 the sequence $\{F_{\pi^n}\}$ weakly converges to F_π.

From Lemma 2.2, (ii) follows easily. □

The continuity of $\mathrm{CVaR}_\gamma(\tilde{\varphi}_\pi)$ is given in the following theorem.

Theorem 2.1 *For any level* γ *with* $0 < \gamma < 1$, $\mathrm{CVaR}_\gamma(\tilde{\varphi}_\pi)$ *is continuous in* $\pi \in \Pi$.

Proof: Since $\underline{M} \leq \tilde{\varphi}_\pi \leq \overline{M}$ a.s-P_π, it holds that $\underline{M} \leq F^{-1}_{-\tilde{\varphi}_\pi}(\gamma) \leq \overline{M}$. So, restricting the range of the infimum, from Proposition 1.1 in Section 1 we have

$$\mathrm{CVaR}_\gamma(\tilde{\varphi}_\pi) = \inf_{\underline{M} \leq b \leq \overline{M}} \left\{ b + \frac{1}{1-\gamma} \int (x-b)^+ dF_\pi \right\}. \tag{6}$$

Here, we apply Proposition 1.2 in Section 1 to prove the theorem.

Let $\mathcal{G} := \{g_b(x) \in C[\underline{M}, \overline{M}] | \underline{M} \leq b \leq \overline{M}\}$, where

$$g_b(x) := b + \frac{1}{1-\gamma}(x-b)^+. \tag{7}$$

Obviously, the class $\mathscr{G}$ is uniformly bounded and equicontinuous at each $x \in [\underline{M}, \overline{M}]$. Writing the equation (6) as $\mathrm{CVaR}_\gamma(\tilde{\varphi}_\pi) = \inf_{g \in \mathcal{G}} \int g_b(x) dF_\pi$, for any $\pi, \pi' \in \Pi$, it yields that

$$|\mathrm{CVaR}_\gamma(\tilde{\varphi}_\pi) - \mathrm{CVaR}_\gamma(\tilde{\varphi}_{\pi'})| \leq \sup_{g \in \mathcal{G}} \left| \int g(x) dF_\pi - \int g(x) dF_{\pi'} \right|. \tag{8}$$

Applying Proposition 1.2 in Section 1 together with Lemma 2.3 (ii), the right hand in (8) converges to 0 as $\pi \to \pi'$, which implies the continuity of $\mathrm{CVaR}_\gamma(\tilde{\varphi}_\pi)$. This completes the proof. □

Now we are ready to state the existence theorem of risk-optimal policies.

Theorem 2.2 *For any $\nu \in P(S)$, there exists a $\pi^* \in \Pi$ which is risk-optimal for ν.*

Proof: By Theorem 2.1, $\mathrm{CVaR}_\gamma(\tilde{\varphi}_\pi)$ is continuous in $\pi \in \Pi$. From Lemma 2.1(i), the class Π is compact. Thus, it follows that there exists a $\pi^* \in \Pi$, for which $\mathrm{CVaR}_\gamma(\tilde{\varphi}_\pi)$ is minimized over all $\pi \in \Pi$. This means that π^* is risk-optimal for ν. □

Corollary 2.1 *There exists a π^* which is risk-optimal for each initial state $i \in S$.*

Proof: Theorem 2.2 shows that for each $i \in S$ there is a $\pi(i)$ which is risk-optimal for i. Here, we define the new policy $\pi^* \in \Pi$ by

$$\pi^* = \pi(i) \ \ if \ \ X_0 = i \ \ (i \in S).$$

Obviously, π^* is risk-optimal for each $i \in S$. □

3 Further Results

In this section, the convex analysis for our optimization problem are given, for which the existence of risk-optimal and deterministic policies will be shown.

Lemma 3.1 $\Phi(\nu)$ *is convex.*

Proof: For any $\pi = (\pi_0, \pi_1, \ldots) \in \Pi$, let $\pi\{n\} := (\pi_0, \pi_1, \ldots, \pi_{n-1})$ $n \geq 1$. Observing that for any $h_n \in \Omega_n$. $P_\pi(H_n = h_n)$ is depending on $\pi\{n\}$, it holds that $P_\pi(H_n = h_n) = P_{\pi\{n\}}(H_n = h_n)$ and $P_{\pi\{n\}}(\cdot) \in P(\Omega_n)$. Let us prove that for any $\pi^1, \pi^2 \in \Pi$ and $0 \leq \alpha \leq 1$, there exists $\pi \in \Pi$ such that

$$P_\pi = \alpha P_{\pi^1} + (1-\alpha) P_{\pi^2} \tag{9}$$

which implies that

$$F_\pi = \alpha F_{\pi^1} + (1-\alpha) F_{\pi^2}. \tag{10}$$

We note that by Kolmogorov extension theorem, for (9) to holds it is sufficient to show that the following equation holds:

$$P_{\pi\{n\}} = \alpha P_{\pi^1\{n\}} + (1-\alpha) P_{\pi^2\{n\}} \ \ for \ \ n \geq 1. \tag{11}$$

The construction of $\pi\{n\}$ satisfying the above equation proceeds inductively.

For $n = 0$, (11) is obviously true.

For the inductive step, assume that (11) hold for n. For $h_n \in \Omega_n$, let $H_{n+1} = (h_n, a, j) \in \Omega_{n+1}$. Here, we define $Q \in P(\Omega_n)$ by

$$\begin{aligned} Q(H_n = h_n, \Delta_n = a_n, X_{n+1} = j) :=& \alpha P_{\pi^1\{n\}}(H_n = h_n) q_{x_n j}(a) \pi^1(a|h_n) \\ &+ (1-\alpha) P_{\pi^2\{n\}}(H_n = h_n) q_{x_n j}(a) \pi^2(a|h_n) \end{aligned} \tag{12}$$

for $h_n \in \Omega_n, a \in A, j \in S$. Then, from the inductive hypothesis, we have

$$\begin{aligned} Q(H_n = h_n) :=& \sum_{a\in A, j\in S} Q(H_n = h_n, \Delta_n = a, X_{n+1} = j) \\ =& \alpha P_{\pi^1\{n\}}(H_n = h_n) + (1-\alpha) P_{\pi^2\{n\}}(H_n = h_n) \\ =& P_{\pi\{n\}}(H_n = h_n), \end{aligned} \tag{13}$$

$$\begin{aligned} Q(H_n = h_n, \Delta_n = a) :=& \sum_{j\in S} Q(H_n = h_n, \Delta_n = a, X_{n+1} = j) \\ =& \alpha P_{\pi^1\{n\}}(H_n = h_n)\pi^1(a|h_n) + (1-\alpha) P_{\pi^2\{n\}}(H_n = h_n)\pi^2(a|h_n). \end{aligned} \tag{14}$$

Here, we define $\pi_n(\cdot|h_n)$ as follows: For each $a \in A$,

$$\pi_n(a|h_n) = \begin{cases} \frac{Q(H_n=h_n, \Delta_n=a)}{Q(H_n=h_n)}, & if \ \ Q(H_n = h_n) > 0 \\ \eta(a), & otherwise \end{cases} \tag{15}$$

where $\eta(\cdot) \in P(A)$ is a arbitrarily given probability.

We put $\pi\{n+1\} = (\pi\{n\}, \pi_n) = (\pi_0, \pi_1, \ldots, \pi_{n-1}, \pi_n)$. Then, for $h_{n+1} = (h_n, a, j) \in \Omega_{n+1}$,

$$\begin{aligned} P_{\pi\{n+1\}}(H_{n+1} = h_{n+1}) &= P_{\pi\{n\}}(H_n = h_n)\pi_n(a|h_n) q_{x_n,j}(a) \\ &= Q(H_n = h_n)\pi_n(a|h_n) q_{x_n,j}(a) \ \ (from\ (13)) \\ &= Q(H_n = h_n, \Delta_n = a) q_{x_n,j}(a) \ \ (from\ (15)) \\ &= Q(H_n = h_n, \Delta_n = a, X_{n+1} = j) \ \ (from\ (12)\ and\ (14)) \\ &= \alpha P_{\pi^1\{n+1\}}(H_{n+1} = h_{n+1}) + (1-\alpha) P_{\pi^2\{n+1\}}(H_{n+1} = h_{n+1}), \end{aligned}$$

which shows that (11) holds for $n+1$. This completes the proof. □

By Lemma 2.3 and 3.1, $\Phi(\nu)$ is convex and weak-compact. The set of extreme points of $\Phi(\nu)$ will be denoted by $Ext\ \Phi(\nu)$. We say that the policy $\pi = (\pi_0, \pi_1, \ldots) \in \Pi$ is deterministic if $\pi(\cdot|h_n)$ is a Dirac measure on A for each $h_n \in \Omega_n$ and $n \geq 0$. We denote by Π_D the set of deterministic policies. The set of distribution functions of $\tilde{\varphi}_\pi$ corresponding to $\pi \in \Pi_D$ is denoted by $\Phi_D(\nu)$, that is

$$\Phi_D(\nu) := \{F_\pi | \pi \in \Pi_D\}.$$

Lemma 3.2 $\Phi_D(\nu) \supset Ext\ \Phi(\nu)$.

Proof: Let $\pi \notin \Pi_D$. To show $F^\pi \notin Ext\ \Phi(\nu)$, we prove that there exists $\pi^1, \pi^2 \in \Pi$ ($\pi^1 \neq \pi^2$) with

$$P_\pi = \frac{1}{2}P_{\pi^1} + \frac{1}{2}P_{\pi^2}. \tag{16}$$

As $\pi = (\pi_0, \pi_1, \ldots) \notin \Pi_D$, there are $n \geq 1$ and $\overline{h}_n \in \Omega_n$ for which it holds that $P_\pi(H_n = \overline{h}_n) > 0$ and $\pi_n(\cdot|\overline{h}_n)$ is not a Dirac measure. So, π_n is represented as

$$\pi_n(\cdot|\overline{h}_n) = \frac{1}{2}\psi^1(\cdot) + \frac{1}{2}\psi^2(\cdot) \tag{17}$$

for some $\psi^1, \psi^2 \in P(A)$ with $\psi^1 \neq \psi^2$. Here, using $\pi = (\pi_0, \pi_1, \ldots), \psi^1$ and ψ^2, we construct $\pi^1 = (\pi^1_0, \pi^1_1, \pi^1_2, \ldots)$, $\pi^2 = (\pi^2_0, \pi^2_1, \pi^2_2, \ldots)$ as follows:

$$\pi^l_t := \pi_t \quad if \quad t \neq n, \quad (t \geq 0) \ \ (l = 1, 2,) \tag{18}$$

and

$$\pi_n^1(\cdot|h_n) = \begin{cases} \psi^1(\cdot) & if \;\; h_n = \overline{h}_n \\ \pi_n(\cdot|h_n) & otherwise, \end{cases}$$
$$\pi_n^2(\cdot|h_n) = \begin{cases} \psi^2(\cdot) & if \;\; h_n = \overline{h}_n \\ \pi_n(\cdot|h_n) & otherwise. \end{cases} \tag{19}$$

Now, we prove that (16) holds for this π^1 and π^2. To this end, by the Kolmogorov extension theorem, it suffices that

$$P_\pi(H_t = h_t) = \frac{1}{2}P_{\pi^1}(H_t = h_t) + \frac{1}{2}P_{\pi^2}(H_t = h_t) \tag{20}$$

for all $h_t \in \Omega_t$ and $t \geq 0$.

In case $0 \leq t \leq n-1$, by (18), (20) obviously holds.

In case $t = n$, by (19) it holds that

$$P_\pi(\Delta_n = a|H_n = h_n) = \begin{cases} \frac{1}{2}\pi_n^1(a|\overline{h}_n) + \frac{1}{2}\pi_n^2(a|\overline{h}_n) & if \;\; h_n = \overline{h}_n \\ \pi_n^1(a|h_n) = \pi_n^2(a|h_n) & otherwise. \end{cases}$$

This shows that (20) holds for $t = n$.

In case $t \geq n+1$, we rewrite $h_t \in \Omega_t$ as $h_t = (h_t^{(n)}, a_n, h_t^{(t|n)})$, where $h_t^{(n)} = (x_0, a_0, \ldots, x_n)$, $h_t^{(t|n)} = (x_{n+1}, a_{n+1}, \ldots, x_t)$. Similarly, we put $H_t^{(n)} = (X_0, \Delta_0, \ldots, X_n)$, and $H_t^{(t|n)} = (X_{n+1}, \Delta_{n+1}, \ldots, X_t)$. Let $h_t \in \Omega_t$. When $h_t^{(n)} = \overline{h}_n$, we have

$$\begin{aligned}
P_\pi(H_t = h_t) &= P_\pi(H_t^{(n)} = \overline{h}_n)P_\pi(\Delta_n = a_n|H_t^{(n)} = \overline{h}_n) \\
&\qquad \times P_\pi(H_t^{(t|n)} = h_t^{(t|n)}|H_t^{(n)} = h_t^{(n)}, \Delta_n = a_n) \\
&= P_\pi(H_t^{(n)} = \overline{h}_n)(\frac{1}{2}\psi^1(a_n) + \frac{1}{2}\psi^2(a_n)) \\
&\qquad \times P_\pi(H_t^{(t|n)} = h_t^{(t|n)}|H_t^{(n)} = h_t^{(n)}, \Delta_n = a_n) \\
&= \frac{1}{2}P_{\pi^1}(H_t^{(n)} = \overline{h}_n)P_{\pi^1}(\Delta_n = a_n|H_t^{(n)} = \overline{h}_n) \\
&\qquad \times P_{\pi^1}(H_t^{(t|n)} = h_t^{(t|n)}|H_t^{(n)} = h_t^{(n)}, \Delta_n = a_n) \\
&\quad + \frac{1}{2}P_{\pi^2}(H_t^{(n)} = \overline{h}_n)P_{\pi^2}(\Delta_n = a_n|H_t^{(n)} = \overline{h}_n) \\
&\qquad \times P_{\pi^2}(H_t^{(t|n)} = h_t^{(t|n)}|H_t^{(n)} = h_t^{(n)}, \Delta_n = a_n) \\
&= \frac{1}{2}P_{\pi^1}(H_t = h_t) + \frac{1}{2}P_{\pi^2}(H_t = h_t).
\end{aligned}$$

When $h_t^{(n)} \neq \overline{h}_n$, by (18) and (19) we have

$$P_\pi(H_t = h_t) = P_{\pi^1}(H_t = h_t) = P_{\pi^2}(H_t = h_t).$$

From the above the proof is complete. □

We need the following lemma, which is obtained by an application of Choquet's theorem (Theorem (Choquet) in [15], p.19-p.20).

Lemma 3.3 *For any $\bar{F} \in \Phi(\nu)$, there is a probability measure μ on Ext $\Phi(\nu)$, such that*

$$\int u(x)d\bar{F} = \int_{Ext\ \Phi(\nu)} d\mu \int u(x)dF \quad for\ all\ u \in C[\underline{M}, \overline{M}].$$

Theorem 3.1 *There exists a deterministic policy which is risk-optimal for ν, that is,*

$$\min_{\pi\in\Pi} \mathrm{CVaR}_\gamma(\tilde{\varphi}_\pi^\nu) = \min_{\pi\in\Pi_D} \mathrm{CVaR}_\gamma(\tilde{\varphi}_\pi^\nu). \tag{21}$$

Proof: From Proposition 1.1 in Section 1, we have

$$\begin{aligned} K &:= \min_{\pi\in\Pi} \mathrm{CVaR}_\gamma(\tilde{\varphi}_\pi) \\ &= \inf_{F\in\Phi(\nu)} \inf_{\underline{M}\le b\le \overline{M}} \int g_b(x)dF(x) \\ &= \inf_{\underline{M}\le b\le \overline{M}} \inf_{F\in\Phi(\nu)} \int g_b(x)dF(x) \end{aligned} \tag{22}$$

where $g_b(x)$ is defined in (7). Applying Lemma 3.3, for any $b \in [\underline{M}, \overline{M}]$ there exists $F^b \in \Phi_D(\nu)$ such that

$$\min_{F\in\Phi(\nu)} \int g_b(x)dF = \int g_b(x)dF^b.$$

By (22) there exists a sequence $\{F^{b_k} \subset \Phi_D(\nu)\}$ such that $\int g_{b_k}(x)dF^{b_k} \to K$ as $k \to \infty$. With the compactness of $\Phi(\nu)$ there is $F^* \in \Phi(\nu)$ for which we can assume without loss of generality, that $\{F^{b_k}\}$ weakly converges to F^*. Since $F^* \in \Phi_D(\nu)$, π^* corresponding F^* is deterministic and risk-optimal for ν. This completes the proof. □

Similarly as Corollary 2.1 in Section 2, we can prove the following.

Corollary 3.1 *There exists a deterministic policy which is risk-optimal for each initial state $i \in S$.*

4 Concluding Remarks

In this paper, we have considered CVaR-minimization problem for finite state MDPs. Using the fact that CVaR of a random income variable depends only on its distribution function, the problem is equivalently redefined as a non-linear programming problem on the attainable set of the distribution functions over all policies, by which the corresponding CVaR-minimization problem is successfully reduced to the one in the area of convex analysis and weak-compactness. Optimal risk or deterministic policies are defined and their existence theorem are proved. The results are purely mathematical. However, the existence theorem is important to develop an algorithm for finding an optimal policy, which will be in future works.

Acknowledgement

We are grateful to Prof. Emeritus M. Kurano and Prof. M. Yasuda of Chiba University for stimulating discussion.

References

[1] Artzner, P., F. Delbaen, J.M. Eber, and D. Heath, Coherent measures of risk, *Mathematical Finance*, vol.9, pp.203–228, 1999.

[2] Billingsley, P., *Convergence of Probability Measures*, Wiley, New York, 1968.

[3] Boda, K., and J.A. Filar, Time consistent dynamic risk measures, *Mathematical Methods of Operations Research*, (Special issue in honor of Arie Hordijk), pp.1–19, 2005.

[4] Borkar, V.S., *Topics in Controlled Markov Chains*, Longman Scientific and Technical, Harlow, 1991.

[5] Derman, C., *Finite State Markov Decision Processes*, Academic Press, New York and London, 1970.

[6] Föllmer, H., and I. Penner, Convex risk measures and the dynamics of their penalty functions, *Statistics & Decisions*, vol.24, pp.61–96, 2006.

[7] Föllmer, H., and A. Schied, Convex measures of risk and trading constraints, *Finance and Stochastics*, vol.6, pp.429–447, 2002.

[8] Föllmer, H., and A. Schied, *Stochastic Finance*, Walter de Gruyter, Berlin and New York, 2004.

[9] Gotoh, J., and Y. Takano, Newsvendor solutions via conditional value-at-risk minimization, *European Journal of Operational Research*, vol.179, pp.80–96, 2007.

[10] Hernandez-Lerma, O., *Adaptive Markov Control Processes*, Springer-Verlag, New York, 1989.

[11] Inoue, A., On the worst conditional expectation, *Journal of Mathematical Analysis and Applications*, vol.286, pp.237–247, 2003.

[12] Kusuoka, S., On law invariant coherent risk measures, *Advances in Mathematical Economics*, vol.3, pp.83–95, 2001.

[13] Mundt, A.P., *Dynamic Risk Management with Markov Decision Processes*, Universitatsverlag Karlsruhe, 2007.

[14] Ohtsubo, Y., and K. Toyonaga, Optimal policy for minimizing risk models in Markov decision processes, *Journal of Mathematical Analysis and Applications*, vol.271, pp.66–81, 2002.

[15] Phelps, R., *Lectures on Choquet's Theorem*, Van Nostrand, 1966.

[16] Puterman, M.L., *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley & Sons Inc, New York, 2004.

[17] Rao, R.R., Relations between weak and uniform convergence of measures with application, *The Annals of Mathematical Statistics*, vol.33, pp.659–680, 1962.

[18] Rockafellar, R.T., *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.

[19] Rockafellar, R.T., and S. Uryasev, Optimization of conditional value-at-risk, *Journal of Risk*, vol.2, pp.21–42, 2000.

[20] Rockafellar, R.T., and S. Uryasev, Conditional value-at-risk for general loss distributions, *Journal of Banking and Finance*, vol.26, pp.1443–1471, 2002.

[21] Takeda, A., Generalization performance of ν-support vector classifier based on conditional value-at-risk minimization, *Neurocomputing*, vol.72, pp.2351–2358, 2009.

[22] White, D.J., Minimizing a threshold probability in discounted Markov decision processes, *Journal of Mathematical Analysis and Applications*, vol.173, pp.634–646, 1993.

[23] Wu, C., and Y. Lin, Minimizing risk models in Markov decision processes with policies depending on target values, *Journal of Mathematical Analysis and Applications*, vol.231, pp.47–67, 2002.