应用概率统计　第 37 卷
第 6 期　2021 年 12 月

Chinese Journal of Applied Probability and Statistics
Dec., 2021, Vol. 37, No. 6, pp. 598-610
doi: 10.3969/j.issn.1001-4268.2021.06.004

可变折扣马氏决策过程首达模型列的收敛问题 *

吴　晓
(肇庆学院数学与统计学院, 肇庆, 526061)

郭圳滨*
(广发证券股份有限公司发展研究中心, 上海, 200120)

摘　要: 本文主要研究了可数状态空间上带多约束、可变折扣马氏决策过程首达模型序列的收敛问题. 利用 "占有测度" 及其相关性质, 将受约束首达模型序列的优化问题转化为等价的受约束线性规划问题 (凸分析方法), 在合适条件下证明了首达模型序列的最优值和最优策略收敛于 "极限" 模型的最优值和最优策略.

关键词: 马氏决策过程首达模型; 多约束; 依赖状态折扣因子; 凸分析方法; 收敛问题

中图分类号: O211.62

英文引用格式: WU X, GUO Z B. Convergence problem of a sequence of first passage Markov decision processes with varying discount factors [J]. Chinese J Appl Probab Statist, 2021, 37(6): 598–610. (in Chinese)

§1.　引　　言

本文主要研究可数状态空间和紧 Borel 行动空间上的带多约束、离散时间马氏决策过程 (简记为 DTMDPs) 首达模型序列的收敛问题, 其折扣因子是依赖于状态的函数, 且成本 (或报酬) 函数是无界的.

常数折扣因子的 DTMDPs 是最常用的模型, 而且其有相当丰富的研究成果. 当状态空间有限时, Puterman [1] 与 Hernández-Lerma 和 Lasserre [2] 在各自不同的条件下证明了折扣最优方程的解以及折扣最优平稳策略的存在性. 对可数状态甚至更一般的 Borel 状态空间的情形, 文献 [3–6] 等都给出了其最优策略存在性条件, 并给出了计算最优策略的方法. 同时, 对常数折扣因子的受约束 DTMDPs 模型的研究, 也得到了很大地发展, 相关研究成果出现在文献 [3, 7]. 后来, 随着凸分析方法 (通过引入占有测度, 将受约束问题转化为一个等价的线性规划问题) 的发展, 对一般状态空间上受约束折扣 DTMDPs, 也有了更多的研究, 如文献 [8–11].

然而, 在经济和金融领域, 折扣因子 (或称贴现因子) 通常是利率的函数, 而利率是与资金数量 (常被看作是系统状态) 和决策者的行动密切相关的. 因而, 讨论与状态和行动有

---

*国家自然科学基金项目 (批准号: 11961005) 和广东省普通高校特色创新类项目基金 (批准号: 2018KTSCX253) 资助.
*通讯作者, E-mail: guozhenbin@gf.com.cn.
本文 2020 年 7 月 26 日收到, 2020 年 9 月 15 日收到修改稿.

关 (或仅与状态有关) 的折扣因子, 在实际应用中显得更有意义. 对于折扣因子依赖状态和行动的 DTMDPs, 主要有 Schäl[12] 等的工作. 但其报酬 (或费用) 函数被要求是有下界或有上界的, 而且, 也没有建立最优方程解的唯一性问题. 对于折扣因子依赖于状态的折扣 DTMDPs, Wei 和 Guo[13] 研究了 (从上和从下) 无界的报酬函数的情形, 在合适条件下, 证明了最优确定性平稳策略的存在性和最优方程解的唯一性, 且给出了计算最优值 (至少是逼近最优值) 的一个算法. 但在上述文献中, 考虑的时间范围大多数都是确定的: 有限或无限. 而在实际应用中, 我们常需要在一个随机时间范围内找出最优策略. 例如: 证券投资优化中, 人们常常需要在到达某个基准资本或者破产前得到最大利益, 这就是我们要研究首达模型的目的. 近年来, 关于可变折扣马氏决策过程的首达模型优化问题的研究, 主要有文献 [14–16].

此外, 对可变折扣因子的受约束 DTMDPs 模型的研究也已取得进展. 例如: 对可数状态空间上的单约束可变折扣因子的 DTMDPs, Huang 等[17] 证明了其首达目标准则的受约束最优策略的存在性. 对多约束情形, Zhang[18] 将凸分析方法发展到了 Borel 状态和行动空间上的有下界成本的受约束 DTMDPs, 且在某些条件下证明了最优策略的存在性; Wu 等[16] 对可数状态空间和紧 Borel 行动空间上的受多约束、无界成本、依赖状态折扣因子 DTMDPs 首达模型, 证明了其最优值和最优随机平稳策略的存在性; 2020 年, Wu 和 Guo[19] 对受约束、可变折扣 DTMDPs 模型, 给出了有限逼近问题的相关结果.

本文将在文献 [16, 19] 的基础上, 进一步研究可数状态空间和紧 Borel 行动空间上的受多约束、无界成本、依赖状态折扣因子 DTMDPs 首达模型的收敛问题, 为进一步的最优值和最优随机平稳策略的有限逼近数值计算提供夯实的理论基础.

§2. 受约束首达模型及首达目标准则

考虑受约束 DTMDPs 首达模型序列 $\{\mathscr{M}_n\}$ $(n\in\overline{\mathbb{N}}:=\mathbb{N}\cup\infty)$ 如下:

$$\mathscr{M}_n:=\big\{S_n,A_n,A_n(i),Q_n(j\,|\,i,a),B,\alpha(i),c_n^0(i,a),(c_n^l(i,a),d_n^l,1\leqslant l\leqslant q),\gamma_n\big\},\qquad(1)$$

其中, S_∞ 为可数状态空间, 且记为 $S_\infty:=\{0,1,2,\cdots\}$; S_n 是 S_∞ 的子集, 且 $\{S_n\}_{n\geqslant1}$ 是单调非减的 (即对每个 $n\geqslant1$ 有 $S_n\subseteq S_{n+1}$) 满足 $S_n\uparrow S$; B 为一给定的目标集; 行动空间 A_n 是 Borel 空间, 其 Borel σ-域为 $\mathscr{B}(A_n)$; 对每个 $i\in S_n$, $A_n(i)$ 为可允许行动集. 对每个 $i\in S_\infty$, 令 $n(i):=\min\{n\geqslant1,\,i\in S_n\}$, 且对 $i\in S_\infty$ 和 $n\geqslant n(i)$, 假定 $A_n(i)$ 是 $A_\infty(i)$ 的可测子集, 故有 $\mathscr{B}(A_n(i))=\mathscr{B}(A_\infty(i))\cap A_n(i)$. 令 $K_n:=\{(i,a)\,|\,i\in S_n,\,a\in A_n(i)\}$ 为所有可允许状态 – 行动对构成的集合, 且假设 K_n 是 $S_n\times A_n$ 的 Borel 可测子集. 转移率 $Q_n(j\,|\,i,a)$ 是给定 $(i,a)\in K_n$ 和 $i,j\in S_n$ 定义在 S_n 上的一步 (时齐) 转移概率. 折扣因子

$\alpha(i)$ 是 S_n 到 $(0,1)$ 上的实值函数. 另外, $c_n^0(i,a)$ 和 $c_n^l(i,a)$ $(1\leqslant l\leqslant q)$ 分别是目标成本和约束成本函数, 且对每个固定的 $i\in S_n$, 其在 $A_n(i)$ 上可测. 最后, 实数 d_n^l $(1\leqslant l\leqslant q)$ 为约束值, 且令 γ_n 是 S_n 上的初始分布.

对每个 $n\in\overline{\mathbb{N}}$, 令 $H_{n,0}:=S_n$, 且对 $m=1,2,\cdots$, $H_{n,m}:=K_n^m\times S_n$. $\mathscr{M}_n$ 的随机历史依赖策略集 Π_n、随机平稳策略集 Φ_n 等的定义可参见文献 [1,2,19,20].

对 S_n 上的任意初始分布 γ_n 和随机历史依赖策略 $\pi=\{\pi_m\}\in\Pi_n$, 由著名的 Tulcea 定理 (见文献 [2; p.178]), 在空间 $(\Omega_n,\mathscr{F}_n)$ 上存在唯一的概率 $\mathsf{P}_{\gamma_n}^\pi$ (见文献 [2; p.16]) 和一个诱导的状态 – 行动随机过程 $\{i_m,a_m,m=0,1,\cdots\}$, 使得对每个 $C\subseteq S_n$, $\Gamma\in\mathscr{B}(A_n)$ 和 $m\geqslant 0$, 有

$$\begin{aligned}
&\mathsf{P}_{\gamma_n}^\pi(i_0\in C)=\gamma_n(C),\\
&\mathsf{P}_{\gamma_n}^\pi(a_m\in\Gamma\,|\,h_m)=\pi_m(\Gamma\,|\,h_m),\\
&\mathsf{P}_{\gamma_n}^\pi(i_{m+1}\in C\,|\,h_m,a_m)=Q_n(C\,|\,i_m,a_m),
\end{aligned}$$

其中, $\mathsf{E}_{\gamma_n}^\pi$ 是 $\mathsf{P}_{\gamma_n}^\pi$ 的期望. 若 γ_n 是集中在某个状态 i 处的 Dirac 测度, 则 $\mathsf{P}_{\gamma_n}^\pi$ 和 $\mathsf{E}_{\gamma_n}^\pi$ 分别记为 $\mathsf{P}_{i,n}^\pi$ 和 $\mathsf{E}_{i,n}^\pi$.

定义 1 对每个 $\pi\in\Pi_n$, $0\leqslant l\leqslant q$ 和初始分布 γ_n, 定义受约束 DTMDPs 的首达目标准则为:

$$\begin{aligned}
V_n^l(\pi)&:=\mathsf{E}_{\gamma_n}^\pi\Big[\sum_{m=0}^{\tau_B-1}\prod_{k=0}^{m-1}\alpha(i_k)c_n^l(i_m,a_m)\Big]\\
&=\mathsf{E}_{\gamma_n}^\pi\Big[\sum_{m=0}^{\infty}I_{\{i_0\in S_n\setminus B,i_1\in S_n\setminus B,\cdots,i_m\in S_n\setminus B\}}\prod_{k=0}^{m-1}\alpha(i_k)c_n^l(i_m,a_m)\Big],
\end{aligned}\tag{2}$$

其中, $I_{\{i\in S_n\setminus B\}}$ 是示性函数, 且 $\tau_B:=\inf\{m\geqslant 0\,|\,i_m\in B\}$ 是状态过程 $\{i_m,m=0,1,\cdots\}$ 首次进入目标集 B 的时刻 (规定 $\inf\emptyset=\infty$). 此外, 若 γ_n 集中在某状态 i 时, $V_n^l(\pi)$ 就记为 $V_n^l(i,\pi)$.

对每个固定的 $n\in\overline{\mathbb{N}}$, 令 $U_n:=\{\pi\in\Pi_n\,|\,V_n^l(\pi)\leqslant d_n^l,\ 1\leqslant l\leqslant q\}$ (通常假定 $U_n\neq\emptyset$, 由下文引理 14 能保证该假定成立) 是控制模型 $\mathscr{M}_n$ 所有可行策略集, 且 $\mathscr{M}_n$ 的最优值为: $V_n^{0*}=\inf\limits_{\pi\in U_n}V_n^0(\pi)$. 因此, (1) 式中的首达模型 $\mathscr{M}_n$ 的受约束首达目标准则优化问题 COP_n (constrained optimality problem) 定义为

$$\mathrm{COP}_n:\ \min_{\pi\in U_n}V_n^0(\pi).$$

而且, 若 $\pi^*\in U_n$ 是在 $\pi\in U_n$ 上使得 $V_n^0(\pi)$ 达到最小的策略, 则称 π^* 是 COP_n 的受约束最优策略. 由于 $c_n^0(i,a)$ 无界, 因而其能用报酬函数来替代, 从而相应的 COP_n 改为在 $\pi\in U_n$ 上使得 $V_n^0(\pi)$ 达到最大值.

条件 2 (a) 存在常数 $\alpha \in (0,1)$ 使得 $\sup\limits_{i \in S_n \setminus B} \alpha(i) \leqslant \alpha$;

(b) 存在非负常数 $\beta \in (0, 1/\alpha)$, $p > 1$ 和 $L > 0$, 以及 $S_n \setminus B$ 上的 Lyapunov 函数 ω (即 ω 是单调不减的, 且 $\lim\limits_{i \to \infty} \omega(i) = \infty$) 满足 $\omega \geqslant 1$, 使得:

$$\sum_{i \in S_n \setminus B} \omega^p(i) \gamma_n(i) < \infty, \qquad \sup_{a \in A_n(i)} |c_n^l(i,a)| \leqslant L\omega(i) \quad (l = 0, 1, \cdots, q),$$

且

$$\sup_{a \in A_n(i)} \sum_{j \in S_n \setminus B} Q_n(j \,|\, i, a) \omega(j) \leqslant \beta \omega(i), \qquad \forall\, i \in S_n \setminus B.$$

(c) 对每个 $i \in S_n \setminus B$, $A_n(i)$ 是紧集;

(d) 对每个 $0 \leqslant l \leqslant q$ 和 $i \in S_n \setminus B$, 成本函数 $c_n^l(i, \cdot)$ 和转移率 $Q_n(j \,|\, i, \cdot)$ $(\forall\, j \in S_n \setminus B)$ 在 $A_n(i)$ 上均连续;

(e) 对每个 $i \in S_n \setminus B$, 函数 $\sum\limits_{j \in S_n \setminus B} Q_n(j \,|\, i, \cdot) \omega(j)$ 在 $A_n(i)$ 上连续.

注记 3 条件 2(c) 说明了: 对每个 $i \in S_n \setminus B$, $\mathsf{P}(A_n(i))$ 按测度弱收敛拓扑的空间是紧的, 其中 $\mathsf{P}(A_n(i))$ 是 $\mathscr{B}(A_n(i))$ 上的所有概率测度之集. 从而, 由 Tychonoff 定理, Φ_n 也是紧的. 另外, 对每个 $i \in S_\infty \setminus B$ 和 $n \geqslant n(i)$, 由于 $A_n(i) \subset A_\infty(i)$, 显然有 $\mathsf{P}(A_n(i)) \subset \mathsf{P}(A_\infty(i))$.

§3. 受约束 DTMDPs 首达模型的收敛结果

本节中, 我们将定义模型序列 $\{\mathscr{M}_n\}$ 的占有测度, 得到占有测度的相关性质与结果, 并以此证明由 (1) 定义的模型列 $\{\mathscr{M}_n\}$ 收敛到 $\{\mathscr{M}_\infty\}$.

条件 4 (a) 对每个 $i, j \in S_\infty \setminus B$ 和 $n \geqslant \max\{n(i), n(j)\}$, 有

$$\lim_{n \to \infty} \sup_{a \in A_n(i)} \left| Q_n(j \,|\, i, a) - Q_\infty(j \,|\, i, a) \right| = 0.$$

(b) 对所有 $i \in S_\infty \setminus B$, $0 \leqslant l \leqslant q$ 和 $n \geqslant n(i)$, 有

$$\lim_{n \to \infty} \sup_{a \in A_n(i)} \left| c_n^l(i, a) - c_\infty^l(i, a) \right| = 0.$$

(c) 对每个 $i \in S_\infty \setminus B$ 和 $n \geqslant n(i)$, 有 $\lim\limits_{n \to \infty} \gamma_n(i) = \gamma_\infty(i)$.

(d) 对每个 $1 \leqslant l \leqslant q$, 有 $\lim\limits_{n \to \infty} d_n^l = d_\infty^l$.

定义 5 对每个 $n\in\overline{\mathbb{N}}$ 和策略 $\pi\in\Pi_n$, 关于 COP_n 的策略 π 的占有测度 $\mu_n^\pi(i,\Gamma)$ 定义如下:

$$\mu_n^\pi(i,\Gamma):=\mathsf{E}_{\gamma_n}^\pi\Big[\sum_{m=0}^{\tau_B-1}\prod_{k=0}^{m-1}\alpha(i_k)I_{\{\{i\}\times\Gamma\}}(i_m,a_m)\Big],\qquad\forall\,\Gamma\in\mathscr{B}(A_n).\tag{3}$$

显然, 由条件 2 (a) 可知, $\mu_n^\pi(i,\Gamma)$ 是 $S_n\backslash B\times A_n$ 上的有限测度. 所有占有测度之集按测度弱收敛拓扑构成的空间记为 $\mathscr{D}_n$, 且 $\mu_n^\pi(i,\Gamma)$ 在 $S_n\backslash B$ 上的边缘测度定义为:

$$\widehat{\mu}_n^\pi(i):=\mu_n^\pi(i,A_n(i)).$$

注记 6 (a) 由于 $\mu_n\in\mathscr{D}_n$ 是定义在 $\widetilde{K}_n:=\{(i,a)\,|\,i\in S_n\backslash B,\,a\in A_n(i)\}=K_n\cap(S_n\backslash B\times A_n)$ 上的, 不失一般性, 以下我们将占有测度 μ_n 看作是 $\widetilde{K}_n$ 上的测度, 因而有

$$V_n^l(\pi)=\sum_{i\in S_n\backslash B}\int_{A_n(i)}c_n^l(i,a)\mu_n^\pi(i,\mathrm{d}a).$$

(b) 让 $\mathscr{M}(\widetilde{K}_n)$ 是定义在 $\widetilde{K}_n$ 上的有限测度集, τ 是弱拓扑, 则 $(\mathscr{M}(\widetilde{K}_n),\tau(\mathscr{M}(\widetilde{K}_n)))$ 是可度量化的, 因而本文将 $(\mathscr{M}(\widetilde{K}_n),\tau(\mathscr{M}(\widetilde{K}_n)))$ 和 $(\mathscr{D}_n,\tau(\mathscr{D}_n))$ 均看作是度量空间.

以下定理 7 和推论 8 与文献 [16] 中的定理 3.4 和推论 3.5 类似, 我们略去证明过程.

定理 7 在条件 2 下, 以下结论成立:

(a) 对每个 $n\in\overline{\mathbb{N}}$ 和 $\pi_n\in\Pi_n$, 占有测度 μ_n^π 满足方程:

$$\mu(i,A_n(i))=\gamma_n(i)+\sum_{j\in S_n\backslash B}\int_{A_n(j)}\alpha(j)Q_n(i\,|\,j,a)\mu(j,\mathrm{d}a)\qquad\forall\,i\in S_n\backslash B,\tag{4}$$

和

$$\sum_{i\in S_n\backslash B}\omega(i)\mu(i,A_n(i))\leqslant\frac{1}{1-\alpha\beta}\sum_{i\in S_n\backslash B}\omega(i)\gamma_n(i)<\infty.\tag{5}$$

(b) 反之, 对每个固定的 $n\in\overline{\mathbb{N}}$, 若 $\widetilde{K}_n$ 上的有限测度 μ 满足 (4)–(5) 式, 则存在随机平稳策略 $\varphi\in\Phi_n$ 使得 $\mu=\mu_n^\varphi$, 也即是, μ 是一占有测度; 且对每个 $i\in S_n\backslash B$ 有 $\widehat{\mu}(i)=\widehat{\mu}_n^\varphi(i)$, 策略 φ 满足分解式: $\mu(i,\mathrm{d}a)=\widehat{\mu}(i)\varphi(\mathrm{d}a\,|\,i)$, 简记为 $\mu=\widehat{\mu}\circ\varphi$.

(c) 对固定的 $n\in\overline{\mathbb{N}}$ 和策略 $\varphi\in\Phi_n$, $(\widehat{\mu}_n^\varphi(i),\,i\in S_n\backslash B)$ 是 $S_n\backslash B$ 上使得 $\sum\limits_{i\in S_n\backslash B}\omega(i)\nu(i)<\infty$ 成立的有限测度类中满足下列方程的唯一解:

$$\nu(i)=\gamma_n(i)+\sum_{j\in S_n\backslash B}\alpha(j)Q_n(i\,|\,j,\varphi)\nu(j).\tag{6}$$

推论 8 假设条件 2 满足, 对固定的 $n\in\overline{\mathbb{N}}$, 以下断言成立.

(a) 占有测度集 $\mathscr{D}_n$ 是凸的.

(b) 对每个策略 $\pi \in \Pi_n$, 存在一 (随机) 平稳策略 $\varphi \in \Phi_n$, 使得对所有 $l = 0, 1, \cdots, q$, $V_n^l(\pi) = V_n^l(\varphi)$. 而且, 若存在 COP_n 的一个最优策略 $\pi \in \Pi_n$, 则必存在 COP_n 的最优平稳策略 $\varphi \in \Phi_n$.

(c) 对每个平稳策略 $\varphi \in \Phi_n$, 存在唯一的占有测度 $\mu \in \mathscr{D}_n$ 使得 $\mu = \widehat{\mu} \circ \varphi$.

注记 9 定理 7 给出了 (3) 中占有测度的判定条件, 且定理 7 (c) 中的唯一性结论将在验证边缘测度相关性质时起到决定性作用.

以下给出占有测度的相关性质.

引理 10 令 $\varphi_n\ (\in \Phi_n)$ 弱收敛到 $\varphi\ (\in \Phi_\infty)$, 则以下结论成立.

(a) 若 $\nu_n\ (\in \mathscr{M}(S_n \backslash B))$ 弱收敛到 $\nu\ (\in \mathscr{M}(S_\infty \backslash B))$, 且有 $\mu_n := \nu_n \circ \varphi_n$ 及 $\mu := \nu \circ \varphi$, 则 μ_n 弱收敛到 μ.

(b) 若条件 2 和 4 成立, 则对每个 $i, j \in S_\infty \backslash B$ 和 $n \geqslant \max\{n(i), n(j)\}$, 有

$$\lim_{n\to\infty} Q_n(j \,|\, i, \varphi_n) = Q_\infty(j \,|\, i, \varphi) \quad 且 \quad \lim_{n\to\infty} c_n^l(i, \varphi_n) = c_\infty^l(i, \varphi), \quad 0 \leqslant l \leqslant q.$$

证明: 由文献 [21] 中的引理 4.6 (a) 可知 (a) 成立, 由文献 [22] 中的引理 5.2 (ii) 的证明可知 (b) 也成立. $\square$

引理 11 在条件 2 和 4 下, 以下结论成立.

(a) 对每个 $n \in \overline{\mathbb{N}}$, 若 $\{\mu_n\}\ (\mu_n \in \mathscr{D}_n)$ 使得 μ_n 弱收敛到 $\mu\ (\in \mathscr{D}_\infty)$ 成立, 则对每个 $l \in \{0, 1, \cdots, q\}$, 有

$$\lim_{n\to\infty} \sum_{i\in S_n\backslash B} \int_{A_n(i)} c_n^l(i, a) \mu_n(i, \mathrm{d}a) = \sum_{i\in S_\infty\backslash B} \int_{A(i)} c_\infty^l(i, a) \mu(i, \mathrm{d}a).$$

(b) 若 $\varphi_n\ (\in \Phi_n)$ 弱收敛到 $\varphi\ (\in \Phi_\infty)$, 则有 $\mu_n^{\varphi_n}$ 弱收敛到 $\mu_\infty^{\varphi}\ (\in \mathscr{D}_\infty)$.

证明: (a) 固定 l, 令

$$f_n := \sum_{i\in S_n\backslash B} \int_{A_n(i)} c_n^l(i, a) \mu_n(i, \mathrm{d}a) \qquad \forall\, n = 0, 1, \cdots,$$

则由条件 2 和 (5) 式可知 $\{f_n\}$ 有界. 对 $\{f_n\}$ 的任意收敛子列 $\{f_{n_k}\}$ (不妨设其收敛到某常数 c), 由定理 7 知: 存在子列 $\{\varphi_{n_k}\}\ (\varphi_{n_k} \in \Phi_{n_k})$ 使得 $\mu_{n_k} = \widehat{\mu}_{n_k} \circ \varphi_{n_k}$. 由 Φ_{n_k} 的紧性并利用对角线方法可知: 存在 $\{\varphi_{n_k}\}$ 的子列 (不失一般性仍记为 $\{\varphi_{n_k}\}$) 使得 φ_{n_k} 弱收敛到 $\varphi\ (\in \Phi_\infty)$. 又由于 μ_{n_k} 弱收敛到 μ, 故对每个 $i \in S \backslash B$ 和 $n_k \geqslant n(i)$, 有 $\widehat{\mu}_{n_k}(i)$ 弱收敛到 $\widehat{\mu}(i)$. 于是, 由引理 10 (a) 可知 $\mu = \widehat{\mu} \circ \varphi$.

对每个 $\widehat{\mu}_{n_k}$ ($k\in\mathbb{N}$), 定义其关联策略 $\widetilde{\mu}_{n_k}$ 如下:

$$\widetilde{\mu}_{n_k}(i):=\begin{cases}\widehat{\mu}_{n_k}(i), & i\in S_{n_k}\backslash B;\\ 0, & \text{其它}.\end{cases}$$

则由条件 2 和 (5) 式可知:

$$\sum_{i\in S_\infty\backslash B}\omega(i)\widehat{\mu}(i)\leqslant\overline{\lim_{k\to\infty}}\sum_{i\in S_{n_k}\backslash B}\omega(i)\widetilde{\mu}_{n_k}(i)\leqslant\frac{1}{1-\alpha\beta}\sum_{i\in S_{n_k}\backslash B}\omega(i)\gamma_{n_k}(i)<\infty.$$

于是, 对任意 $\varepsilon>0$, 存在正整数 N 使得:

$$\sum_{i\in S_\infty\backslash B,\, i\geqslant N}\omega(i)\widehat{\mu}(i)\leqslant\frac{\varepsilon}{3},$$

且有

$$\begin{aligned}\sum_{i\in S_\infty\backslash B,\, i\geqslant N}\omega(i)\widetilde{\mu}_{n_k}(i)&=\sum_{i\in S_{n_k}\backslash B,\, i\geqslant N}\omega(i)\widehat{\mu}_{n_k}(i)\\&\leqslant\frac{1}{\omega^{p-1}(N)}\sum_{i\in S_{n_k}\backslash B,\, i\geqslant N}\omega^p(i)\widehat{\mu}_{n_k}(i)\\&\leqslant\frac{1}{\omega^{p-1}(N)}\cdot\frac{1}{1-\alpha\beta}\sum_{i\in S_{n_k}\backslash B,\, i\geqslant N}\omega^p(i)\gamma_{n_k}(i)\\&\leqslant\frac{\varepsilon}{3}.\end{aligned}$$

又因为 $\widehat{\mu}_{n_k}(i)\to\widehat{\mu}(i)$, 则对每个 $i\in S_\infty\backslash B$ 和 $i<N$, 存在正整数 $M(i)$ 使得当 $k\geqslant M(i)$ 时, 有

$$|\widetilde{\mu}_{n_k}(i)-\widehat{\mu}(i)|\leqslant\frac{\varepsilon}{3N\omega(i)}.$$

因此, 对每个 $k\geqslant\max\{M(i):i\leqslant N-1\}$, 我们有

$$\begin{aligned}&\Big|\sum_{i\in S_{n_k}\backslash B}\omega(i)\widehat{\mu}_{n_k}(i)-\sum_{i\in S_\infty\backslash B}\omega(i)\widehat{\mu}(i)\Big|\\=&\Big|\sum_{i\in S_\infty\backslash B}\omega(i)\widetilde{\mu}_{n_k}(i)-\sum_{i\in S_\infty\backslash B}\omega(i)\widehat{\mu}(i)\Big|\\=&\Big|\sum_{i\in S_\infty\backslash B,\, i<N}\omega(i)[\widetilde{\mu}_{n_k}(i)-\widehat{\mu}(i)]+\sum_{i\in S_\infty\backslash B,\, i\geqslant N}\omega(i)[\widetilde{\mu}_{n_k}(i)-\widehat{\mu}(i)]\Big|\\\leqslant&\Big|\sum_{i=0}^{N}\omega(i)[\widetilde{\mu}_{n_k}(i)-\widehat{\mu}(i)]\Big|+\frac{2\varepsilon}{3}\leqslant\varepsilon.\end{aligned}$$

于是有

$$\sum_{i\in S_\infty\backslash B}\omega(i)\widehat{\mu}(i)=\lim_{k\to\infty}\sum_{i\in S_{n_k}\backslash B}\omega(i)\widehat{\mu}_{n_k}(i)<\infty.$$

注意到, 对每个 $0 \leqslant l \leqslant q$, 由引理 10 (b) 可知 $c^l_{n_k}(i, \varphi_{n_k}) \to c^l_\infty(i, \varphi)$. 因此, 由推广的控制收敛定理 (见文献 [23; Theorem A.2.6]) 可得:

$$\sum_{i \in S_{n_k} \setminus B} c^l_{n_k}(i, \varphi_{n_k}) \widehat{\mu}_{n_k}(i) \to \sum_{i \in S_\infty \setminus B} c^l_\infty(i, \varphi) \widehat{\mu}(i),$$

即

$$f_{n_k} = \sum_{i \in S_{n_k} \setminus B} \int_{A(i)} c^l(i, a) \mu_{n_k}(i, \mathrm{d}a) \to \sum_{i \in S_\infty \setminus B} \int_{A(i)} c^l(i, a) \mu(i, \mathrm{d}a).$$

由于 $\{f_{n_k}\}$ 是任意选取的, 故有

$$f_n = \sum_{i \in S_n \setminus B} \int_{A(i)} c^l(i, a) \mu_n(i, \mathrm{d}a) \to \sum_{i \in S_\infty \setminus B} \int_{A(i)} c^l(i, a) \mu(i, \mathrm{d}a) = c, \quad \forall\, l = 0, 1, \cdots, q.$$

(b) 由引理 10 (a) 可知, 我们只需证明 $\widehat{\mu}^{\varphi_n}_n$ 弱收敛到 $\widehat{\mu}^{\varphi}_\infty$ 即可. 任取 $\{\widehat{\mu}^{\varphi_n}_n\}$ 的子列 $\{\widehat{\mu}^{\varphi_{n_k}}_{n_k}\}$, 设其收敛到某个 $S_\infty \backslash B$ 上的测度 ν. 则, 对每个 $j \in S_\infty \backslash B$ 和 $n_k \geqslant \max\{n(i), n(j)\}$, 由定理 7 (c), 我们有

$$\widehat{\mu}^{\varphi_{n_k}}_{n_k}(i) = \gamma_{n_k}(i) + \sum_{j \in S_{n_k} \setminus B} \alpha(j) Q_{n_k}(i \,|\, j, \varphi_{n_k}) \widehat{\mu}^{\varphi_{n_k}}_{n_k}(j). \tag{7}$$

另一方面, 由 (a) 的证明过程, 我们有

$$\sum_{j \in S_{n_k} \setminus B} \alpha(j) Q_{n_k}(i \,|\, j, \varphi_{n_k}) \widehat{\mu}^{\varphi_{n_k}}_{n_k}(j) = \sum_{j \in S_\infty \setminus B} \alpha(j) Q_\infty(i \,|\, j, \varphi) \nu(j),$$

联立 (7) 可得:

$$\nu(i) = \gamma_\infty(i) + \sum_{j \in S_\infty \setminus B} \alpha(j) Q_\infty(i \,|\, j, \varphi) \nu(j).$$

因此, 由定理 7 (c) 可知: 对所有 $i \in S_\infty \backslash B$, 有 $\nu(i) = \widehat{\mu}^{\varphi}_\infty(i)$ 成立. $\square$

对每个 $n \in \overline{\mathbb{N}}$, 记

$$\boldsymbol{F}_n := \left\{ \mu_n \in \mathscr{D}_n \,\middle|\, \sum_{i \in S_n \setminus B} \int_{A_n(i)} c^l_n(i, a) \mu_n(i, \mathrm{d}a) \leqslant d^l_n,\ 1 \leqslant l \leqslant q \right\},$$

则带约束优化问题 COP_n 等价于下述凸线性规划问题:

$$\mathrm{COP}'_n: \quad \min_{\mu_n \in \boldsymbol{F}_n} \sum_{i \in S_n \setminus B} \int_{A_n(i)} c^0_n(i, a) \mu_n(i, \mathrm{d}a).$$

$\boldsymbol{F}_n$ 称为是 COP'_n 的所有可行解之集, 且 COP'_n 的最优值记为 $\inf \mathrm{COP}'_n$. 若存在 COP'_n 的一个可行解 $\mu \in \boldsymbol{F}_n$ 使得

$$\sum_{i \in S_n \setminus B} \int_{A_n(i)} c^0_n(i, a) \mu(i, \mathrm{d}a) = \inf \mathrm{COP}'_n,$$

则 μ 称为是 COP'_n 的最优解, 且所有最优解之集记为 $\boldsymbol{O}_n$. 由假设 $U_n \neq \emptyset$ 可知 $\boldsymbol{F}_n \neq \emptyset$.

而且, 对任意 $\mu \in \boldsymbol{F}_n$, 由条件 2 (b) 和 (5) 式可知

$$\begin{aligned}\sum_{i\in S_n\backslash B}\int_{A_n(i)} c_n^0(i,a)\mu(i,\mathrm{d}a) &\leqslant L\sum_{i\in S_n\backslash B}\int_{A_n(i)}\omega(i)\mu(i,\mathrm{d}a) = L\sum_{i\in S_n\backslash B}\omega(i)\widehat{\mu}(i)\\ &\leqslant \frac{L}{1-\alpha\beta}\sum_{i\in S_n\backslash B}\omega(i)\gamma_n(i) < \infty,\end{aligned}\tag{8}$$

故有 $\inf \mathrm{COP}'_n < \infty$.

定理 12 若条件 2 和 4 成立, 则对每个 $n \in \overline{\mathbb{N}}$, 有:

(a) 占有测度空间 $\mathscr{D}_n$ 和 COP'_n 的可行解 $\boldsymbol{F}_n$ 按测度弱收敛拓扑构成的空间都是紧空间.

(b) 令 $\mu_n \in \boldsymbol{F}_n$, 则 $\{\mu_n\}$ 在 $\boldsymbol{F}_\infty$ 中是相对紧的, 且 $\{\mu_n\}$ 的任一聚点都属于 $\boldsymbol{F}_\infty$.

证明: (a) 由文献 [16] 中的引理 3.9 和 4.1 可知 (a) 成立.

(b) 令 $\{\mu_{n_k}\}$ 是 $\{\mu_n\}$ 的任一子列. 则由定理 7 (b) 知, 存在 $\{\varphi_{n_k}\} \in \Phi_{n_k}$ 使得 $\mu_{n_k} = \mu_{n_k}^{\varphi_{n_k}}$. 而且, 由 Φ_n 的紧性, 存在 $\{\varphi_{n_k}(\cdot\,|\,i)\}$ 的子列 (不妨仍记为 $\{\varphi_{n_k}(\cdot\,|\,i)\}$) 和 $\varphi(\cdot\,|\,i) \in A_\infty(i)$, 使得 $\{\varphi_{n_k}(\cdot\,|\,i)\}$ 弱收敛到 $\varphi(\cdot\,|\,i)$. 由于 $S\backslash B$ 是可数集, 由对角线方法知: 存在子列 $\{\varphi_{n_{k_s}}\}$ 使得 $\{\varphi_{n_{k_s}}\}$ 弱收敛到 $\varphi \in \Phi_\infty$. 联立引理 11 (b) 可得 $\mu_{n_{k_s}} = \mu_{n_{k_s}}^{\varphi_{n_{k_s}}}$ 弱收敛到 μ_∞^φ. 而且, 由引理 11 (a) 可知:

$$\begin{aligned}\sum_{i\in S_\infty\backslash B}\int_{A_\infty(i)} c_\infty^l(i,a)\mu_\infty^\varphi(i,\mathrm{d}a) &= \lim_{s\to\infty}\sum_{i\in S_{n_{k_s}}\backslash B}\int_{A_{n_{k_s}}(i)} c_{n_{k_s}}^l(i,a)\mu_{n_{k_s}}^{\varphi_{n_{k_s}}}(i,\mathrm{d}a)\\ &\leqslant \lim_{s\to\infty} d_{n_{k_s}}^l = d_\infty^l, \qquad \forall\, 1\leqslant l\leqslant q.\end{aligned}$$

故有 $\mu_\infty^\varphi \in \boldsymbol{F}_\infty$, 因而结论 (b) 成立. □

条件 13 (Slater 条件) 存在策略 $\pi\in\Pi_\infty$, 使得对所有 $1\leqslant l\leqslant q$, 有 $V_\infty^l(\pi)<d_\infty^l$ 成立.

下面的引理类似于文献 [22; 命题 5.1], 我们略去证明过程.

引理 14 若条件 2、4 和 13 成立. 则存在正整数 N_1、常数 $\delta > 0$ 和策略 $\widetilde{\varphi}_n \in \Phi_n$, 使得

$$V_n^l(\widetilde{\varphi}_n) \leqslant d_n^l - \delta, \qquad \forall\, 1\leqslant l\leqslant q,\ n\geqslant N_1.$$

即, 对所有 $n \geqslant N_1$, $U_n \neq \emptyset$.

定理 15 若条件2、4和13成立, 则有以下结论:

(a) 存在正整数 N_2, 使得对所有 $n \geqslant N_2$, 有 $\boldsymbol{F}_n \neq \emptyset$.

(b) 对每个 $\mu \in \boldsymbol{F}_\infty$, 存在正整数 N_3 和 $\mu_n \in \boldsymbol{F}_n$ $(n \geqslant N_3)$ 使得 μ_n 弱收敛到 μ.

(c) 令 $\mu_n \in \boldsymbol{O}_n$ $(n \geqslant 1)$, 则 $\{\mu_n\}$ 的任一聚点都属于 $\boldsymbol{O}_\infty$.

证明: (a) 由引理 14, 结论 (a) 显然成立.

(b) 令 $\mu \in \boldsymbol{F}_\infty$, 于是存在 Φ_∞ 中策略 φ^1 与 φ^2, 使得 $\mu = \mu_\infty^{\varphi^1}$, 且对所有 $1 \leqslant l \leqslant q$, 存在常数 $C > 0$ 使得 $V_\infty^l(\varphi^2) \leqslant d_\infty^l - C$. 对每个 $n \geqslant 1$, $i \in S_n \backslash B$ 和 φ^k $(k = 1, 2)$, 可定义 φ_n^k $(k = 1, 2)$ 如下

$$\varphi_n^k(\mathrm{d}a \,|\, i) := \begin{cases} \varphi^k(\mathrm{d}a \cap A_n(i) \,|\, i)/\varphi^k(A_n(i) \,|\, i), & \varphi^k(A_n(i) \,|\, i) > 0; \\ I_{\{a_n(i)\}}(\mathrm{d}a), & \varphi^k(A_n(i) \,|\, i) = 0, \end{cases}$$

其中 $a_n(i) \in A_n(i)$ 是任意给定的. 由于对 $n \geqslant n(i)$ 有 $A_n(i) \uparrow A_\infty(i)$, 故有 φ_n^k 弱收敛到 φ^k $(k = 1, 2)$ 且 $\varphi^k \in \Phi_\infty$. 于是, 由引理 11 (b) 可得 $\mu_n^{\varphi_n^k} \to \mu_\infty^{\varphi^k}$. 因而, 对所有 $1 \leqslant l \leqslant q$, 由引理 11 (a) 我们有

$$\lim_{n\to\infty} V_n^l(\varphi_n^1) = V_\infty^l(\varphi^1) \leqslant d_\infty^l,$$
$$\lim_{n\to\infty} V_n^l(\varphi_n^2) = V_\infty^l(\varphi^2) \leqslant d_\infty^l - C.$$

因此, 对任意固定的 $0 < \varepsilon < C/4$, 存在正整数 $N(\varepsilon)$, 使得对每个 $1 \leqslant l \leqslant q$ 和 $n \geqslant N(\varepsilon)$, 有

$$|d_n^l - d_\infty^l| \leqslant \varepsilon, \quad V_n^l(\varphi_n^1) \leqslant d_\infty^l + \varepsilon, \quad \text{且} \quad V_n^l(\varphi_n^2) \leqslant d_\infty^l - C + \varepsilon.$$

令 $\nu_n := (1 - \theta_\varepsilon)\mu_n^{\varphi_n^1} + \theta_\varepsilon \mu_n^{\varphi_n^2}$, 其中 $\theta_\varepsilon := 2\varepsilon/C < 1/2$. 则, 对每个 $1 \leqslant l \leqslant q$ 和 $n \geqslant N(\varepsilon)$, 我们有

$$\begin{aligned} \sum_{i\in S_n\backslash B} \int_{A_n(i)} c_n^l(i, a)\nu_n(i, \mathrm{d}a) &= (1 - \theta_\varepsilon)V_n^l(\varphi_n^1) + \theta_\varepsilon V_n^l(\varphi_n^2) \\ &\leqslant (1 - \theta_\varepsilon)(d_\infty^l + \varepsilon) + \theta_\varepsilon(d_\infty^l - C + \varepsilon) = d_n^l. \end{aligned}$$

于是, 对所有 $n \geqslant N(\varepsilon)$, 有 $\nu_n \in \boldsymbol{F}_n$.

进一步, 对每个 $s \geqslant 3$, 取 $\varepsilon_s := C/(8s)$. 由上述讨论可知, 对每个 $s \geqslant 3$, 存在正整数 N_s (不失一般性, 假设 N_s 在 $s \geqslant 3$ 上非减) 使得:

$$\nu_n^s := (1 - \theta_s)\mu_n^{\varphi_n^1} + \theta_s \mu_n^{\varphi_n^2} \in \boldsymbol{F}_n, \qquad \forall\, n \geqslant N_s,$$

其中 $\theta_s := 2\varepsilon_s/C$. 因而, 对每个 $N_s \leqslant n < N_{s+1}$, 可取 $\mu_n := \nu_n^s$, 则对 $n \geqslant N_3$, 有 $\mu_n \in \boldsymbol{F}_n$. 而且, 由于 $\varepsilon_s \downarrow 0$, 故有 $\mu_n \to \mu_\infty^{\varphi^1} = \mu$. 所以, 结论 (b) 成立.

(c) 令 μ_∞ 是 $\{\mu_n\}$ 的任一聚点, 即, 存在 $\{\mu_n\}$ 的子列 $\{\mu_{n_k}\}$ 使得 μ_{n_k} 弱收敛到 μ_∞. 由定理 12 (b), 显然有 $\mu_\infty \in \boldsymbol{F}_\infty$. 另一方面, 对任意 $\mu \in \boldsymbol{F}_\infty$, 由 (b) 可知: 存在一正整数 N 和 $\mu'_n \in \boldsymbol{F}_n$, 使得对所有 $n \geqslant N$, 有 μ'_n 弱收敛到 μ. 由于 $\mu_{n_k} \in \boldsymbol{O}_{n_k}$, 故

$$\sum_{i\in S_{n_k}\backslash B} \int_{A_{n_k}(i)} c_{n_k}^0(i, a)\mu_{n_k}(i, \mathrm{d}a) \leqslant \sum_{i\in S_{n_k}\backslash B} \int_{A_{n_k}(i)} c_{n_k}^0(i, a)\mu'_{n_k}(i, \mathrm{d}a) \qquad \forall\, n_k \geqslant N,$$

结合引理 11 (a) 可得:

$$\sum_{i\in S_\infty\backslash B}\int_{A_\infty(i)} c_\infty^0(i,a)\mu_\infty(i,\mathrm{d}a)\leqslant \sum_{i\in S_\infty\backslash B}\int_{A_\infty(i)} c_\infty^0(i,a)\mu(i,\mathrm{d}a).$$

即是, $\mu_\infty\in\boldsymbol{O}_\infty$. □

下面给出收敛问题的主要结果:

定理 16 若条件 2、4 和 13 成立, 则以下结论成立.

(a) 若 V_∞^{0*} 与 V_n^{0*} 分别是 $\mathscr{M}_\infty$ 与 $\mathscr{M}_n$ 的最优值, 则有 $\lim\limits_{n\to\infty}V_n^{0*}=V_\infty^{0*}$.

(b) 若 π_n^* $(n\geqslant 1)$ 是 $\mathscr{M}_n$ 的最优策略, 则 $\{\pi_n^*\}$ 的任意聚点均是 $\mathscr{M}_\infty$ 的最优策略.

证明: (a) 令 $\{V_{n_k}^{0*}\}$ 是 $\{V_n^{0*}\}$ 的任意收敛子列, 则存在子列 $\{\mu_{n_k}^*\}$ $(\mu_{n_k}^*\in\boldsymbol{O}_{n_k})$ 使得:

$$V_{n_k}^{0*}=\sum_{i\in S_{n_k}\backslash B}\int_{A_{n_k}(i)} c_{n_k}^0(i,a)\mu_{n_k}^*(i,\mathrm{d}a).$$

又由定理 12 (b) 和 15 (c), 存在 $\{\mu_{n_k}^*\}$ 的子列 (不妨仍记为本身) 使得当 $k\to\infty$ 时 $\mu_{n_k}^*$ 弱收敛到 μ_∞^*, 且有 $\mu_\infty^*\in\boldsymbol{O}_\infty$. 联立引理 11 (a) 可得:

$$\begin{aligned}\lim_{n\to\infty}V_{n_k}^{0*}&=\lim_{n\to\infty}\sum_{i\in S_{n_k}\backslash B}\int_{A_{n_k}(i)} c_{n_k}^0(i,a)\mu_{n_k}^*(i,\mathrm{d}a)\\&=\sum_{i\in S_\infty\backslash B}\int_{A_\infty(i)} c_\infty^0(i,a)\mu_\infty^*(i,\mathrm{d}a)=V_\infty^{0*}.\end{aligned}$$

因此, 由 $\{V_{n_k}^0*\}$ 的任意性可知结论 (a) 成立.

(b) 注意到: 对每个 $n\in\overline{\mathbb{N}}$, 策略 π_n^* 是 $\mathscr{M}_n$ 的最优策略当且仅当其相应的占有测度 $\mu_n^{\pi_n^*}\in\boldsymbol{O}_n$. 令 π_n^* $(n\geqslant 1)$ 是 $\mathscr{M}_n$ 的最优策略, 且 π_∞ 是 $\{\pi_n^*\}$ 的任一聚点, 即, 存在 $\{\pi_n^*\}$ 的子列 $\{\pi_{n_k}^*\}$ 使得 $\pi_{n_k}^*$ 弱收敛到 π_∞ (不失一般性, 由推论 8 我们可假定 $\pi_n^*\in\Phi_n$ 且 $\pi_\infty\in\Phi_\infty$). 则由引理 11 (b) 可得: $\mu_{n_k}^{\pi_{n_k}^*}$ 弱收敛到 $\mu_\infty^{\pi_\infty}$. 因为 $\mu_{n_k}^{\pi_{n_k}^*}\in\boldsymbol{O}_{n_k}$, 则由定理 15 (c) 可得 $\mu_\infty^{\pi_\infty}\in\boldsymbol{O}_\infty$. 因而, π_∞ 是 $\mathscr{M}_\infty$ 的最优策略. 故结论 (b) 成立. □

参 考 文 献

[1] PUTERMAN M L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming* [M]. New York: Wiley, 1994.

[2] HERNÁNDEZ-LERMA O, LASSERRE J B. *Discrete-Time Markov Control Processes: Basic Optimality Criteria* [M]. New York: Springer-Verlag, 1996.

[3] ALTMAN E. *Constrained Markov Decision Processes* [M]. Florida: Chapman and Hall/CRC, 1999.

[4] HERNÁNDEZ-LERMA O, LASSERRE J B. *Further Topics on Discrete-Time Markov Control Processes* [M]. New York: Springer-Verlag, 1999.

[5] HERNÁNDEZ-LERMA O, HERNÁNDEZ-HERNÁNDEZ D. Discounted cost Markov decision processes on Borel spaces: the linear programming formulation [J]. *J Math Anal Appl*, 1994, **183(2)**: 335–351.

[6] PIUNOVSKIY A B. *Optimal Control of Random Sequences in Problems with Constraints* [M]. Dordrecht: Kluwer Academic, 1997.

[7] ALTMAN E. Denumerable constrained Markov decision processes and finite approximations [J]. *Math Oper Res*, 1994, **19(1)**: 169–191.

[8] GONZÁLEZ-HERNÁNDEZ J, HERNÁNDEZ-LERMA O. Extreme points of sets of randomized strategies in constrained optimization and control problems [J]. *SIAM J Optim*, 2005, **15(4)**: 1085–1104.

[9] HERNÁNDEZ-LERMA O, GONZÁLEZ-HERNÁNDEZ J. Constrained Markov control processes in Borel spaces: the discounted case [J]. *Math Methods Oper Res*, 2000, **52(2)**: 271–285.

[10] MAO X, PIUNOVSKIY A. Strategic measures in optimal control problems for stochastic sequences [J]. *Stochastic Anal Appl*, 2000, **18(5)**: 755–776.

[11] PIUNOVSKIY A B. Controlled random sequences: methods of convex analysis and problems with functional constraints [J]. *Russian Math Surveys*, 1998, **53(6)**: 1233–1293.

[12] SCHÄL M. Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal [J]. *Z Wahrscheinlichkeitstheorie Verw Gebiete*, 1975, **32(3)**: 179–196.

[13] WEI Q D, GUO X P. Markov decision processes with state-dependent discount factors and unbounded rewards/costs [J]. *Oper Res Lett*, 2011, **39(5)**: 369–374.

[14] WU X, GUO X P. First passage optimality and variance minimisation of Markov decision processes with varying discount factors [J]. *J Appl Probab*, 2015, **52(2)**: 441–456.

[15] WU X, ZHANG J Y. Finite approximation of the first passage models for discrete-time Markov decision processes with varying discount factors [J]. *Discrete Event Dyn Syst*, 2016, **26(4)**: 669–683.

[16] WU X, ZOU X L, GUO X P. First passage Markov decision processes with constraints and varying discount factors [J]. *Front Math China*, 2015, **10(4)**: 1005–1023.

[17] HUANG Y H, WEI Q D, GUO X P. Constrained Markov decision processes with first passage criteria [J]. *Ann Oper Res*, 2013, **206**: 197–219.

[18] ZHANG Y. Convex analytic approach to constrained discounted Markov decision processes with non-constant discount factors [J]. *TOP*, 2013, **21(2)**: 378–408.

[19] WU X, GUO X P. Convergence of Markov decision processes with constraints and state-action dependent discount factors [J]. *Sci China Math*, 2020, **63(1)**: 167–182.

[20] GUO X P, HERNÁNDEZ-LERMA O. *Continuous-Time Markov Decision Processes: Theory and Applications* [M]. Berlin: Springer-Verlag, 2009.

[21] ALVAREZ-MENA J, HERNÁNDEZ-LERMA O. Convergence of the optimal values of constrained Markov control processes [J]. *Math Methods Oper Res*, 2002, **55(3)**: 461–484.

[22] GUO X P, ZHANG W Z. Convergence of controlled models and finite-state approximation for discounted continuous-time Markov decision processes with constraints [J]. *European J Oper Res*, 2014, **238(2)**: 486–496.

[23] SENNOTT L I. *Stochastic Dynamic Programming and the Control of Queueing Systems* [M]. New York: Wiley, 1999.

Convergence Problem of a Sequence of First Passage Markov Decision Processes with Varying Discount Factors

WU Xiao

(*School of Mathematics and Statistics, Zhaoqing University, Zhaoqing, 526061, China*)

GUO Zhenbin

(*Development Research Center, GF Securities Co., Ltd., Shanghai, 200120, China*)

Abstract: In this paper, we study the convergence problem of a sequence of first passage Markov decision processes with constraints and varying discount factors. Using the "occupation measures" and its related properties, we transform the constrained optimality problems into linear programming problems on the set of occupation measures (i.e., the convex analytic approach), and then prove that the optimal values and optimal policies of the original first passage Markov decision processes converge respectively to those of the "limit" one.

Keywords: first passage Markov decision processes; multiple constraints; state-dependent discount factors; convex analytic approach; convergence problem

2020 Mathematics Subject Classification: 60J10; 90C40; 93E20