Realizable strategies in continuous-time Markov decision processes

Alexey Piunovskiy
Department of Mathematical Sciences, University of Liverpool, L69 7ZL, UK.
piunov@liv.ac.uk

Abstract

For the Borel model of the continuous-time Markov decision process, we introduce a wide class of control strategies. In particular case, such strategies transform to the standard relaxed strategies, intensively studied in the last decade. In another special case, if one restricts to another special subclass of the general strategies, the model transforms to the semi-Markov decision process. Further, we show that the relaxed strategies are not realizable. For the constrained optimal control problem with total expected costs, we describe the sufficient class of realizable strategies, the so called Poisson-related strategies. Finally, we show that, for solving the formulated optimal control problems, one can use all the tools developed earlier for the classical discrete-time Markov decision processes.

Keywords: continuous-time Markov decision process, total cost, discounted cost, relaxed strategy, randomized strategy

AMS 2000 subject classification: Primary 90C40; Secondary 60J25, 60J75.

1 Introduction

The theory of continuous-time jump Markov processes is a well developed area of Operational Research, with plenty of fruitful applications: see, e.g., the recent monographs [11, 24]. In the framework of Queueing Theory, the state of the controlled process $X(\cdot)$ can be the number of customers in the system, and the actions can affect the service rate or the intensity of the input stream of the customers. In any case, the controlled process $X(\cdot)$ is assumed to be piece-wise constant, with values in a fixed Borel space $\mathbf{X}$. After the initial state $X(0) = x_0 \in \mathbf{X}$, which can be random, becomes known, the decision maker has to choose the control on the interval $(0, T_1]$, up to the next jump moment T_1.

1. If he/she applies a specific action $a = \varphi(x_0) \in \mathbf{A}$ which can certainly depend on x_0, then the sojourn time $\Theta_1 = T_1$ and the new state $X(T_1) = X_1$ are random:

$$P(\Theta_1 \le t | X(0) = x_0) = 1 - e^{-q_{x_0}(\varphi(x_0))t}, \tag{1}$$

where $q_{x_0}(a)$ is the total jumps intensity, and, in case $q_{x_0}(a) > 0$, for $x_1 \neq x_0$,

$$P(X_1 = x_1 | X(0) = x_0) = \frac{q(\{x_1\}|x_0, \varphi(x_0))}{q_{x_0}(\varphi(x_0))}. \tag{2}$$

Here and below, we assume for simplicity that the state space $\mathbf{X}$ is countable, $\mathbf{A}$ is a fixed standard Borel space of actions, and we use the standard notation $q(\{x_1\}|x_0, a)$ for the jumps intensity; $q_{x_0}(a) = q(\mathbf{X} \setminus \{x_0\}|x_0, a) = -q(\{x_0\}|x_0, a)$. More general and formal definitions are given in the next section. Here, we only underline that formulae (1) and (2) can be combined together, if $q_{x_0}(a) > 0$:

$$P(\Theta_1 \le t,\ X_1 = x_1 | X(0) = x_0) = \int_{(0,t]} q(\{x_1\}|x_0, \varphi(x_0)) e^{-q_{x_0}(\varphi(x_0))s} ds. \tag{3}$$

2. The action $a \in \mathbf{A}$, being similar to 1, can be randomized. Namely, the decision maker can choose the probability space $(\Xi, \mathcal{B}(\Xi), p)$, where Ξ is a standard Borel space, and the probability $p(d\xi|x_0)$ can depend on x_0. After that, the corresponding random element Ξ is simulated taking value ξ, and the action $\varphi(x_0, \xi)$ is applied, where $\varphi: \mathbf{X} \times \Xi \to \mathbf{A}$ is a chosen measurable mapping. Usually, random elements are denoted by capital letters, the lower case is for their realized values.

   If only such actions are allowed, we deal with the so called exponential semi-Markov decision process [25, Ch.7]. Clearly, formula (3) takes the form

$$P(\Theta_1 \le t,\ X_1 = x_1|X(0) = x_0) = \int_{\Xi} \left(\int_{(0,t]} q(\{x_1\}|x_0, \varphi(x_0, \xi)) e^{-q_{x_0}(\varphi(x_0,\xi))s} ds \right) p(d\xi|x_0). \tag{4}$$

   Here again we assume that $x_1 \neq x_0$ and $q_{x_0}(a) > 0$ for all possible actions a. Obviously, if $\Xi = \{1\}$ is a singleton, then we are in the framework of case 1. On another hand, if, e.g., one plans to mix two actions a_1 and a_2 independently of x_0, then $\Xi = \{1, 2\}$, $p(1|x_0) = p(2|x_0) = \frac{1}{2}$ and $\varphi(x_0, i) = a_i$.

   Of course, without loss of generality, here one can take $\Xi = \mathbf{A}$ and put $\varphi(x_0, a) = a$, but in more general situations the space Ξ can be different. Since any uncountable standard Borel space is isomorphic to the segment $[0, 1]$ [3, Co.7.16.1], one can always take $\Xi = [0, 1]$, but again it is convenient to keep the introduced notations.

3. More generally, one can apply different actions depending on time, e.g., according to a measurable mapping $\varphi: \mathbf{X} \times (0, \infty) \to \mathbf{A}$. The randomized version is defined by the measurable mapping $\varphi(x_0, \xi, t)$ from $\mathbf{X} \times \Xi \times (0, \infty)$ to $\mathbf{A}$. If $q_{x_0}(a) > 0$, then expression (4) takes the form

$$\begin{aligned} & P(\Theta_1 \le t,\ X_1 = x_1|X(0) = x_0) \\ = & \int_{\Xi} \left(\int_{(0,t]} q(\{x_1\}|x_0, \varphi(x_0, \xi, s)) e^{-\int_{(0,s]} q_{x_0}(\varphi(x_0,\xi,u))du} ds \right) p(d\xi|x_0). \end{aligned} \tag{5}$$

   Note that the space $(\Xi, \mathcal{B}(\Xi), p)$ can be rather complicated, e.g., in case $\mathbf{A} = \mathbb{R}$, under a fixed x_0, the stochastic process φ defined on $\Xi \times (0, \infty)$ can be a Brownian motion [22, p.3510].

In each of the described situations, after the initial state x_0 becomes known, we have a (complete) probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ and a measurable (w.r.t. $(\tilde{\omega}, t)$) process $A(\cdot)$ on $\tilde{\Omega} \times (0, \infty)$ with values in $\mathbf{A}$. We will call such strategies realizable. In the previous examples, $\tilde{\Omega} = \Xi$, $\tilde{P}(\cdot) = p(\cdot|x_0)$, $\tilde{\mathcal{F}}$ is the completion of $\mathcal{B}(\Xi)$, and $A(\cdot) = \varphi(x_0, \cdot)$.

On the interval $(t_1, T_2]$, after the values $T_1 = t_1$ and $X_1 = X(T_1) = x_1$ become known, the situation is similar. The only difference is that the actions can also depend on t_1 and x_1. And so on.

First continuous-time Markov decision processes were introduced in the 50-60-ies [2, 15], where only the deterministic strategies of the type 1 were considered. More general models were investigated in the 70-ies [26, 28] within the class of deterministic past-dependent strategies. In all the mentioned works, the strategies were realizable and the control process $A(\cdot)$ was well defined.

Starting from the 80-ies [18], the following new class of strategies came to the stage.

4. The decision maker chooses the time-dependent probability distribution $\pi(\cdot|x_0, t)$ on $\mathbf{A}$, which can also depend on x_0. The sojourn time and the new state have the following distribution (again assuming that $x_1 \neq x_0$ and $q_{x_0}(a) > 0$):

$$\begin{aligned} & P(\Theta_1 \le t,\ X_1 = x_1|X(0) = x_0) \\ = & \int_{(0,t]} \int_{\mathbf{A}} q(\{x_1\}|x_0, a)\pi(da|x_0, s) e^{-\int_{0,s]} \int_{\mathbf{A}} q_{x_0}(a)\pi(da|x_0,u))du} ds. \end{aligned} \tag{6}$$

Let us compare the case of x_0-independent stationary (time-independent) $\pi(\cdot)$ and formula (4) with $\mathbf{\Xi} = \mathbf{A}$, x_0-independent measure p, and $\varphi(x_0, a) = a$, taking the form

$$P(\Theta_1 \leq t,\ X_1 = x_1 | X(0) = x_0) = \int_{\mathbf{A}} \left(\int_{(0,t]} q(\{x_1\}|x_0, a) e^{-q_{x_0}(a)s} ds \right) p(da).$$

In the latter case, the sojourn time Θ_1 is not exponential if the p measure is not degenerate, while, according to (6), Θ_1 is exponential with parameter $\int_{\mathbf{A}} q_{x_0}(a)\pi(da)$.

Expression (6) means that "randomizations" are applied independently at any time moment $s \in (0, \infty)$. To distinguish from the case 2, we call such π-strategies "relaxed" rather than randomized. They became very popular in the last decade: see, e.g., [10, 11, 12, 13, 20, 24] and the references therein. In Section 3 we explain what it means that the strategy defined in terms of π can be equivalently represented as a random process $A(\cdot)$. If such a measurable process exists, we say that the strategy is realizable. The main result of the current article states that only the strategies with degenerate kernels π are realizable.

Usually the solutions to constrained optimization problems and to Markov games are given by relaxed strategies [11, Th.11.4], [12, Th.7.1], [13, Th.5.1], [20, Th.3.11], [24, Th.8.6,10.8,10.11]. Another sufficient class of strategies are "mixtures" [12, Th.7.2], [13, Th.5.2], [20, Cor.3.14]. Intuitively, a mixture means that the decision maker flips a coin at the very beginning and afterwards applies this or that deterministic (Markov or stationary) strategy. Such a way to control the process is easy for implementation, but, formally speaking, it cannot be described as a relaxed strategy and does not fit the definition of a strategy introduced in the cited works. Since the relaxed strategies are not realizable if the π kernels are not degenerate (i.e., are different from the Dirac measures), after obtaining a solution to an optimal control problem in terms of a π-strategy, one has to explain how practitioners can use it.

In the case of discounted model, realizable solutions in the form of switching and randomized strategies were constructed in [8, 9]. But if the discount factor α is zero, standard randomized strategies are not sufficient for solving optimal control problems, as demonstrated in Section 4.

In Section 2, we describe the model and the wide class of control strategies including the discussed cases 1-4 and their combinations, along with mixtures. In Section 3 we prove the main results about the realizability of the strategies. Definition 3 seems natural to introduce the concept of the realizability. Theorem 2 states that a strategy is realizable if and only if there is a random process equivalently representing it. Thus, existence of a process satisfying all the assertions of Definition 4 can be accepted as another natural definition of the realizability. Finally, for the constrained models with the total expected cost, we present in Section 4 the sufficient class of realizable strategies, that is, Poisson-related strategies, and show in Section 5 how the tools developed for the discrete-time models can be used for solving continuous-time problems. Note that we investigate the undiscounted Borel model with arbitrarily unbounded transition and cost rates, with the possibility of explosion and with an arbitrary, not necessarily finite number of constraints. All this makes the current article different from the similar works in the area.

The following notations are frequently used throughout this paper. $\mathbb{N} = \{1, 2, \ldots\}$ is the set of natural numbers; $\delta_x(\cdot)$ is the Dirac measure concentrated at x, we call such distributions degenerate; $I\{\cdot\}$ is the indicator function. $\mathcal{B}(E)$ is the Borel σ-algebra of the Borel space E, $\mathcal{P}(E)$ is the Borel space of probability measures on E. (It is always clear which σ-algebra is fixed in E.) The Borel σ-algebra $\mathcal{B}(\mathcal{P}(E))$ comes from the weak convergence of measures, after we fix a proper topology in E. $\mathbb{R}_+ \stackrel{\triangle}{=} (0, \infty)$, $\mathbb{R}_+^0 \stackrel{\triangle}{=} [0, \infty)$, $\bar{\mathbb{R}} = [-\infty, +\infty]$, $\bar{\mathbb{R}}_+ = (0, \infty]$, $\bar{\mathbb{R}}_+^0 = [0, \infty]$; in $\mathbb{R}_+$ and $\mathbb{R}_+^0$, we consider the Borel σ-algebra, and Leb is the Lebesgue measure. The abbreviation *w.r.t.* (resp. *a.s.*) stands for "with respect to" (resp. "almost surely"); for $b \in \bar{\mathbb{R}}$, $b^+ \stackrel{\triangle}{=} \max\{b, 0\}$ and $b^- \stackrel{\triangle}{=} \min\{b, 0\}$. Measures introduced in the current article can take infinite values. Let $(\Omega, \mathcal{F})$ be some measurable space. $\mathcal{G}_1 \vee \mathcal{G}_2$ is the minimal σ-algebra containing the two given σ-algebras $\mathcal{G}_1$ and $\mathcal{G}_2$ in Ω; $\mathcal{F}(X)$ is the σ-algebra generated by a measurable mapping $X:\ \Omega \to \mathbf{X}$, where $(\mathbf{X}, \mathcal{B})$ is another measurable space.

2 Model description and preliminaries

The primitives of a continuous-time Markov decision process are the following elements.

(i) State and action spaces $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$ and $(\mathbf{A}, \mathcal{B}(\mathbf{A}))$ (arbitrary standard Borel).

(ii) Transition rate $q(dy|x,a)$ is a signed kernel on $\mathbf{X}$ given $(x,a) \in \mathbf{X} \times \mathbf{A}$ taking nonnegative values on $\Gamma_{\mathbf{X}} \setminus \{x\}$, where $\Gamma_{\mathbf{X}} \in \mathcal{B}(\mathbf{X})$. We assume that q is conservative in the sense that $q(\mathbf{X}|x,a) = 0$, i.e., $q_x(a) \stackrel{\triangle}{=} q(\mathbf{X}\setminus\{x\}|x,a) = -q(\{x\}|x,a)$. We also assume that the transition rate q is stable, that is, $\sup_{a\in\mathbf{A}} q_x(a) < \infty$ for each $x \in \mathbf{X}$.

(iii) Cost rates $c_n^j(\cdot)$ ($j \in J \cup \{0\}$, $n = 1, 2, \ldots$) are measurable functions on $\mathbf{X} \times \mathbf{A}$ with values in the extended real line $[-\infty, \infty]$; $J \not\ni 0$ is an arbitrary set of indices. Index 0 corresponds to the main objective, the given real numbers d^j, $j \in J$ are the maximal allowed values for other objectives: see problem (15).

(iv) Initial distribution $\gamma(\cdot)$, a probability measure on $\mathbf{X}$.

We need to introduce immediately the standard Borel space $(\Xi, \mathcal{B}(\Xi))$, the source of the control randomness which in fact is chosen by the decision maker, as described in Introduction. We introduce the artificial isolated point (cemetery) Δ, put $\mathbf{X}_\Delta \stackrel{\triangle}{=} \mathbf{X} \cup \{\Delta\}$, $\Xi_\Delta = \Xi \cup \{\Delta\}$, and define $q(\Gamma|\Delta, a) \stackrel{\triangle}{=} 0$ for all $\Gamma \in \mathcal{B}(\mathbf{X}_\Delta)$, $a \in \mathbf{A}$.

Given the above primitives, let us construct the underlying (measurable) sample space $(\Omega, \mathcal{F})$. Having firstly defined the measurable space $(\Omega^0, \mathcal{F}^0) \stackrel{\triangle}{=} (\Xi \times (\mathbf{X} \times \Xi \times \mathbb{R}_+)^\infty, \mathcal{B}(\Xi \times (\mathbf{X} \times \Xi \times \mathbb{R}_+)^\infty))$, let us adjoin all the sequences of the form

$$(\xi_0, x_0,\ \xi_1,\ \theta_1,\ x_1,\ \xi_2,\ \ldots,\ \theta_{m-1},\ x_{m-1},\ \xi_m,\ \infty,\ \Delta,\ \Delta,\ \infty,\ \Delta,\ \Delta,\ \ldots)$$

to Ω^0, where $m \ge 1$ is some integer, $\xi_m \in \Xi$, $\theta_l \in \mathbb{R}_+$, $x_l \in \mathbf{X}$, $\xi_l \in \Xi$ for all nonnegative integers $l \le m-1$. After the corresponding modification of the σ-algebra $\mathcal{F}^0$, we obtain the basic sample space $(\Omega, \mathcal{F})$.

Below,

$$\omega = (\xi_0,\ x_0,\ \xi_1,\ \theta_1,\ x_1,\ \xi_2,\ \theta_2,\ x_2, \ldots).$$

For $n \in \mathbb{N}$, introduce the mapping $\Theta_n : \ \Omega \to \bar{\mathbb{R}}_+$ by $\Theta_n(\omega) = \theta_n$; for $n \in \mathbb{N} \cup \{0\}$, the mappings $X_n : \ \Omega \to \mathbf{X}_\Delta$ and $\Xi_n : \ \Omega \to \Xi_\Delta$ are defined by $X_n(\omega) = x_n$ and $\Xi_n(\omega) = \xi_n$. As usual, the argument ω will be often omitted. The increasing sequence of random variables T_n, $n \in \mathbb{N} \cup \{0\}$ is defined by $T_n = \sum_{i=1}^n \Theta_i$; $T_\infty = \lim_{n\to\infty} T_n$. Here, Θ_n (resp. T_n, X_n) can be understood as the sojourn times (resp. the jump moments, the states of the process on the intervals $[T_n, T_{n+1})$). The realized values of Θ_n, T_n and X_n will be denoted as θ_n, t_n and x_n. We do not intend to consider the process after T_∞. The meaning of the ξ_n components will be described later; see also Introduction. Finally, for $n \in \mathbb{N} \cup \{0\}$,

$$H_n = (\Xi_0, X_0, \Xi_1, \Theta_1, X_1, \ldots, \Xi_n, \Theta_n, X_n)$$

is the n-term (random) history and $\mathbf{H}_n = \{(\xi_0, x_0, \xi_1, \theta_1, x_1, \ldots, \xi_n, \theta_n, x_n)\}$ is the space of all such histories. The controlled process of our interest is

$$X(\omega, t) \stackrel{\triangle}{=} \sum_{n\ge 0} I\{T_n \le t < T_{n+1}\} X_n + I\{T_\infty \le t\}\Delta.$$

Definition 1 *A control strategy is defined as follows:*

$$S = \{\Xi, p_0, \langle p_n, \pi_n \rangle,\ n = 1, 2, \ldots\},$$

where $p_0(d\xi_0)$ is a probability distribution on Ξ; for $x_{n-1} \in \mathbf{X}$, $p_n(d\xi_n|h_{n-1})$ is a stochastic kernel on Ξ given $\mathbf{H}_{n-1}$; $\pi_n(da|h_{n-1}, \xi_n, u)$ is a stochastic kernel on $\mathbf{A}$ given $\mathbf{H}_{n-1} \times \Xi \times \mathbb{R}_+$. If $x_{n-1} = \Delta$, then we assume that $p_n(d\xi_n|h_{n-1}) = \delta_\Delta(d\xi_n)$; the kernels $\pi_n(da|h_{n-1}, \Delta, u)$ are of no importance and can be defined arbitrarily. The set of all control strategies is denoted as Π.

Control on the interval $(T_{n-1}, T_n]$ is based on the both kernels p_n and π_n which respectively correspond to randomizations and relaxations. We underline, that the random element Ξ_n is generated only once, at the jump epoch T_{n-1}. On the opposite, relaxations mean, roughly speaking, that the actions are simulated at each time moment $t = T_{n-1} + u$, continuously in time. For example, the purely randomized Poisson-related strategies discussed in Sections 4 and 5 mean that, at each jump epoch T_{n-1}, the decision maker plans in advance, at which discrete moments in the future the actions will change. This plan may be not deterministic, but, once realized, it does not change as time goes on; hence the kernels p_n do not depend on u.

If the randomizations are absent, that is, the kernels π_n do not depend on the ξ-components, then we deal with a relaxed strategy. One can take $\boldsymbol{\Xi} = \{\tilde{\xi}\}$ as a singleton and simply omit the ξ_n components; as a result we obtain the standard control strategy or policy $\{\pi_n,\ n = 1, 2, \dots\}$ [10, 11, 12, 13, 18, 20, 24] which is called below as a π-strategy. On the other hand, if the relaxations are absent, that is, all kernels π_n are degenerate and concentrated at singletons

$$\varphi_n(\xi_0, x_0, \xi_1, \theta_1, \dots, x_{n-1}, \xi_n, u) \in \mathbf{A}, \tag{7}$$

then we deal with a randomized strategy denoted below as a ξ-strategy. If φ_n does not depend on ξ_0 and u, and one is restricted to such control strategies, then in fact he(she) deals with a semi-Markov decision process (cf. case 2 in Introduction). General control strategies will be sometimes called π-ξ-strategies. The ξ_0 component is responsible for the mixtures of strategies: if, for example, p_0 is a combination of two Dirac measures, then in the future, depending on the realized value ξ_0, this or that control strategy will be used. Clearly, if functions φ_n in (7) do not depend on the ξ-components, then the strategy is purely deterministic and can be equally regarded as a ξ- or π-strategy. A deterministic strategy is called Markov if $\varphi_n(\cdot) = \varphi^M(x_{n-1}, T_{n-1} + u)$; it is called stationary if $\varphi_n(\cdot) = \varphi^S(x_{n-1})$. A more detailed discussion of π-ξ-strategies can be found in [22]. Here, we only underline that on the interval $(t_{n-1}, T_n] \subset \mathbb{R}_+$, $n \in \mathbb{N}$, the jumps intensity is

$$\lambda_n^q(\Gamma_{\mathbf{X}}|h_{n-1}, \xi_n, \theta) = \int_{\mathbf{A}} \pi_n(da|h_{n-1}, \xi_n, \theta) q(\Gamma_{\mathbf{X}} \setminus \{x_{n-1}\}|x_{n-1}, a), \tag{8}$$

where $\Gamma_{\mathbf{X}} \in \mathcal{B}(\mathbf{X})$, $\theta > 0$ is the time elapsed after the realized jump epoch t_{n-1} and ξ_n is the realization of the random element Ξ_n having the distribution $p_n(d\xi_n|h_{n-1})$; h_{n-1} is the realized history with $t_{n-1} = \sum_{i=1}^{n-1} \theta_i < \infty$, $x_{n-1} \in \mathbf{X}$. Along with the intensity λ_n^q, we need the following integral

$$\Lambda_n^q(h_{n-1}, \xi_n, \theta) = \int_{(0,\theta] \cap \mathbb{R}_+} \lambda_n^q(\mathbf{X}|h_{n-1}, \xi_n, u) du. \tag{9}$$

Now the joint distribution of (Θ_n, X_n) is defined by

$$\int_{\Gamma_{\mathbb{R}} \cap \mathbb{R}_+} G_n^{\xi,q}(\Gamma_{\mathbf{X}}, h_{n-1}, \xi_n, \theta) d\theta + I\{\infty \in \Gamma_{\mathbb{R}}\} I\{\Delta \in \Gamma_{\mathbf{X}}\} e^{-\Lambda_n^q(h_{n-1}, \xi_n, \infty)}, \tag{10}$$

where $\Gamma_{\mathbb{R}} \in \mathcal{B}(\bar{\mathbb{R}}_+)$, $\Gamma_{\mathbf{X}} \in \mathcal{B}(\mathbf{X}_\Delta)$, and

$$G_n^{\xi,q}(\Gamma_{\mathbf{X}}, h_{n-1}, \xi_n, \theta) = \lambda_n^q(\Gamma_{\mathbf{X}} \setminus \{\Delta\}|h_{n-1}, \xi_n, \theta) e^{-\Lambda_n^q(h_{n-1}, \xi_n, \theta)}. \tag{11}$$

Under a fixed control strategy S, the probability measure P_γ^S on $(\Omega, \mathcal{F})$, called strategical measure, is build in the standard way. The distribution of $H_0 = (\Xi_0, X_0)$ is given by $p_0(d\xi_0) \cdot \gamma(dx_0)$ and, for any $n \in \mathbb{N}$, the stochastic kernel G_n on $\boldsymbol{\Xi}_\Delta \times \bar{\mathbb{R}}_+ \times \mathbf{X}_\Delta$ given $\mathbf{H}_{n-1}$ is defined by formulae

$$\begin{aligned}
G_n(\boldsymbol{\Xi}_\Delta \times \{\infty\} \times \mathbf{X}|h_{n-1}) &= G_n(\{\Delta\} \times \mathbb{R}_+ \times \mathbf{X}_\Delta|h_{n-1}) = 0. \\
G_n(\{\Delta\} \times \{\infty\} \times \{\Delta\}|h_{n-1}) &= \delta_{x_{n-1}}(\{\Delta\}); \qquad\qquad (12) \\
G_n(\Gamma_{\boldsymbol{\Xi}} \times \{\infty\} \times \{\Delta\}|h_{n-1}) &= \delta_{x_{n-1}}(\mathbf{X}) \int_{\Gamma_{\boldsymbol{\Xi}}} e^{-\Lambda_n^q(h_{n-1}, \xi_n, \infty)} p_n(d\xi_n|h_{n-1}); \\
G_n(\Gamma_{\boldsymbol{\Xi}} \times \Gamma_{\mathbb{R}} \times \Gamma_{\mathbf{X}}|h_{n-1}) &= \delta_{x_{n-1}}(\mathbf{X}) \int_{\Gamma_{\boldsymbol{\Xi}}} \int_{\Gamma_{\mathbb{R}}} G_n^{\xi,q}(\Gamma_{\mathbf{X}}, h_{n-1}, \xi_n, \theta) d\theta \; p_n(d\xi_n|h_{n-1}),
\end{aligned}$$

where $\Gamma_{\mathbf{\Xi}} \in \mathcal{B}(\mathbf{\Xi})$, $\Gamma_{\mathbb{R}} \in \mathcal{B}(\mathbb{R}_+)$, $\Gamma_{\mathbf{X}} \in \mathcal{B}(\mathbf{X}_\Delta)$. It remains to apply the induction and Ionescu Tulcea's theorem [3, Prop.7.28] to obtain P^S_γ. Expectation with respect to P^S_γ is denoted as E^S_γ.

After the history h_{n-1} with $t_{n-1} < \infty$ becomes known, the decision maker flips a coin resulting in the $\Xi_n = \xi_n$ component having distribution $p_n(d\xi_n|h_{n-1})$. After that the stochastic kernel $\pi_n(da|h_{n-1}, \xi_n, u)$ gives rise to the jumps intensity $\underline{\lambda}_n(\Gamma|h_{n-1}, \xi_n, \theta)$ from the current state x_{n-1} to $\Gamma_{\mathbf{X}} \in \mathcal{B}(\mathbf{X})$. After that, the sojourn time $\theta_n \in \bar{\mathbb{R}}_+$ and the new state $x_n \in \mathbf{X}_\Delta$ of the process $X(t)$ at the jump epoch $t_n = t_{n-1} + \theta_n$ are realized according to the joint distribution given by (10). And so on. If the standard Borel spaces $(\Xi_n, \mathcal{B}(\Xi_n))$ are different then one should introduce their direct product $\mathbf{\Xi} = \prod_{n=1}^\infty \Xi_n$.

Definition 2 *Two strategies S^1 and S^2 are called indistinguishable if the space $\mathbf{\Xi}$ is the same, $p_0(\cdot)$ is the common distribution on $\mathbf{\Xi}$, and for each $n \in \mathbb{N}$ the following assertions are valid.*

(a) For $P^{S^1}_\gamma$-almost all H_{n-1} (equivalently, for $P^{S^2}_\gamma$-almost all H_{n-1}), such that $T_{n-1} \neq \infty$, $p^1_n(\cdot|H_{n-1}) = p^2_n(\cdot|H_{n-1})$.

(b) For almost all $u \in \mathbb{R}_+$, $\pi^1_n(\cdot|H_{n-1}, \Xi_n, u) = \pi^2_n(\cdot|H_{n-1}, \Xi_n, u)$ for $P^{S^1}_\gamma$-almost all H_{n-1}, Ξ_n (equivalently, for $P^{S^2}_\gamma$-almost all H_{n-1}, Ξ_n) such that $T_{n-1} \neq \infty$.

For indistinguishable strategies, $P^{S^1}_\gamma = P^{S^2}_\gamma$. Moreover, the detailed occupation measures on $\mathbf{X} \times \mathbf{A}$

$$\eta^S_n(\Gamma_{\mathbf{X}} \times \Gamma_{\mathbf{A}}) = E^S_\gamma \left[\int_{(T_{n-1}, T_n] \cap \mathbb{R}_+} I\{X_{n-1} \in \Gamma_{\mathbf{X}}\} \pi_n(\Gamma_{\mathbf{A}}|H_{n-1}, \Xi_n, t - T_{n-1}) dt \right], \quad n = 1, 2, \ldots, \tag{13}$$

which will be used below to define the objective functionals, coincide for all $n \in \mathbb{N}$, too.

The constructed mathematical model covers both the traditional continuous-time Markov decision processes and exponential semi-Markov decision processes. It makes the base for the investigation of the controlled jump processes without switching between different models like in [8].

For a given cost rate $c^j_n(\cdot)$ on $\mathbf{X} \times \mathbf{A}$, the corresponding objective is defined as

$$\begin{aligned} W^j(S) &= E^S_\gamma \left[\sum_{n=1}^\infty \int_{(T_{n-1}, T_n] \cap \mathbb{R}_+} \int_{\mathbf{A}} \pi_n(da|H_{n-1}, \Xi_n, t - T_{n-1}) c^{j+}_n(X_{n-1}, a) dt \right] \\ &\quad + E^S_\gamma \left[\sum_{n=1}^\infty \int_{(T_{n-1}, T_n] \cap \mathbb{R}_+} \int_{\mathbf{A}} \pi_n(da|H_{n-1}, \Xi_n, t - T_{n-1}) c^{j-}_n(X_{n-1}, a) dt \right]. \end{aligned} \tag{14}$$

Here and below, $\infty - \infty \overset{\triangle}{=} +\infty$; all integrals and series are calculated separately for the positive and negative parts. The constrained optimal control problem under study looks as follows:

$$W^0(S) \to \inf_{S \in \Pi} \text{ subject to } W^j(S) \le d^j, \quad j \in J. \tag{15}$$

In terms of the detailed occupation measures (13), this problem can be rewritten as

$$\begin{aligned} W^0(S) = \sum_{n=1}^\infty \int_{\mathbf{X} \times \mathbf{A}} c^0_n(x, a) \eta^S_n(dx, da) &\quad \to \quad \inf_{S \in \Pi} \\ \text{subject to } W^j(S) = \sum_{n=1}^\infty \int_{\mathbf{X} \times \mathbf{A}} c^j_n(x, a) \eta^S_n(dx, da) &\quad \le \quad d^j, \quad j \in J. \end{aligned} \tag{16}$$

In many works, the admissible sets of actions $\mathbf{A}(x)$ in the states $x \in \mathbf{X}$ were introduced [10, 11, 12, 13, 18, 20, 24, 26, 27, 28]. In this connection, one can, e.g., put $c^0_n(x, a) = +\infty$ in case action a is not admissible in the state x.

3 Realizable strategies

In this section, we present the main results. Intuitively, a strategy is realizable (or implementable) if the actions $A(t)$, to be applied at the time moments $t \in \mathbb{R}_+$, form a measurable random process. A very simple example in [17, Ex.1.2.5] shows that for a relaxed strategy such a process does not exist. Below, we prove that only randomized strategies are realizable. But firstly, let us discuss informally the possible definition of the realizability.

The definition of a strategy is based on the knowledge of the spaces $\mathbf{X}$ and $\mathbf{A}$ only. We plan to construct the definition of realizability, which will be independent of the transition rate and cost rates, as well. Assume, a control strategy $S = \{\Xi, p_0, \langle p_n, \pi_n\rangle,\ n \in \mathbb{N}\}$ is chosen and suppose the pair $(h_{n-1}, \xi_n) \in \mathbf{H}_{n-1} \times \Xi$ is fixed (realized) for a fixed $n \in \mathbb{N}$, and $t_{n-1} < \infty$, $x_{n-1} \in \mathbf{X}$. The actions $A(t)$ for $t > t_{n-1}$ can certainly depend on n, h_{n-1} and ξ_n, as well as on the time $u = t - t_{n-1}$ elapsed after the last jump moment. Below, we omit n, h_{n-1} and ξ_n and formulate the following natural requirements to the process $A(\cdot)$.

Definition 3 *A control strategy S is called realizable for $(h_{n-1}, \xi_n) \in \mathbf{H}_{n-1} \times \Xi$ ($n \in \mathbb{N}$) on the interval $(t_{n-1}, T_n] \neq \emptyset$ if there is a complete probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ and a measurable (with respect to $(u, \tilde{\omega})$) process $A(\cdot)$ on $\mathbb{R}_+ \times \tilde{\Omega}$ with values in $\mathbf{A}$ such that the following properties are satisfied (as usual, the argument $\tilde{\omega} \in \tilde{\Omega}$ will be often omitted):*

(a) $\pi_n(\Gamma_{\mathbf{A}}|h_{n-1}, \xi_n, u)$ coincides with $\tilde{P}(A(u) \in \Gamma_{\mathbf{A}})$ for each $\Gamma_{\mathbf{A}} \in \mathcal{B}(\mathbf{A})$, for almost all $u \in \mathbb{R}_+$.

(b) For any measurable total transition rate $\hat{q}(\cdot)$ on $\mathbf{A}$, the random probability measure $\tilde{G}_{\tilde{\omega}}$ on $\bar{\mathbb{R}}_+$, depending on $\tilde{\omega} \in \tilde{\Omega}$ and defined by

$$\begin{aligned} \tilde{G}_{\tilde{\omega}}(\Gamma_{\mathbb{R}}) &= \int_{\Gamma_{\mathbb{R}} \cap \mathbb{R}_+} \hat{q}(A(\theta, \tilde{\omega})) e^{-\int_{(0,\theta]} \hat{q}(A(u,\tilde{\omega}))du} d\theta, \\ &\quad + I\{\infty \in \Gamma_{\mathbb{R}}\} e^{-\int_{(0,\infty)} \hat{q}(A(u,\tilde{\omega}))du}, \quad \Gamma_{\mathbb{R}} \in \mathcal{B}(\bar{\mathbb{R}}_+), \end{aligned}$$

after taking expectation $\tilde{E}$ with respect to $\tilde{P}$, coincides with the probability measure (10) on $\bar{\mathbb{R}}_+$ at $\Gamma_{\mathbf{X}} = \mathbf{X}_\Delta$, $q_{x_{n-1}}(a) = \hat{q}(a)$.

A control strategy S is called realizable if it is realizable for each $n \in \mathbb{N}$, on $(T_{n-1}, T_n] \neq \emptyset$ for P^S_γ-almost all (H_{n-1}, Ξ_n). The probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ can be different for different n, h_{n-1}, ξ_n.

The requirement $(T_{n-1}, T_n] \neq \emptyset$ is equivalent to $T_{n-1} < \infty$. In all other cases, the length of the interval $(T_{n-1}, T_n]$ is a continuous positive random variable (with a possible atom at ∞), and $X_{n-1} \in \mathbf{X}$ P^S_γ-a.s. As was mentioned, the $A(\cdot)$ process can depend on n, h_{n-1} and ξ_n, but at the moment we do not require the measurability of $A_n(h_{n-1}, \xi_n, u, \tilde{\omega})$ in all the arguments.

Remark 1 *If a strategy S is realizable, then any strategy S', indistinguishable from it, is also realizable.*

Theorem 1 *Suppose the pair $(h_{n-1}, \xi_n) \in \mathbf{H}_{n-1} \times \Xi$ is fixed for some $n \in \mathbb{N}$, such that $t_{n-1} < \infty$. Then the following statements are equivalent*

- *A control strategy S is realizable for (h_{n-1}, ξ_n) on the interval $(t_{n-1}, T_n] \neq \emptyset$.*
- *For almost all $u \in \mathbb{R}_+$, $\pi_n(\cdot|h_{n-1}, \xi_n, u) = \delta_{\varphi(u)}(\cdot)$ is a Dirac measure, where $\varphi(\cdot)$ is an $\mathbf{A}$-valued measurable function on $\mathbb{R}_+$.*

The *proof* is postponed to Appendix.

Corollary 1 *(a) Suppose a control strategy S is realizable for $(h_{n-1}, \xi_n) \in \mathbf{H}_{n-1} \times \Xi$ ($n \in \mathbb{N}$) on the interval $(t_{n-1}, T_n] \neq \emptyset$. Then the process $A(\cdot)$ is in fact non-random in the sense that $\tilde{P}(A(u) = \varphi(u)) = 1$ for almost all $u \in \mathbb{R}_+$, where the $\mathbf{A}$-valued function $\varphi(\cdot)$ is non-random.*

(b) A control strategy S is realizable if and only if it is indistinguishable from a randomized strategy S', that is, a ξ-strategy defined by functions $\varphi_n(\cdot)$ (7), $n \in \mathbb{N}$, and the components $p_0(\cdot), p_1(\cdot), \dots$ coming from the initial strategy S.

Proof. Item (a) immediately follows from Theorem 1.

(b) Suppose a control strategy S is realizable. For a fixed $n \in \mathbb{N}$, the set

$$\{(h_{n-1}, \xi_n, u) \in \mathbf{H}_{n-1} \times \mathbf{\Xi} \times \mathbb{R}_+ : \ \pi_n(\cdot|h_{n-1}, \xi_n, u) \text{ is a Dirac measure}\}$$

is measurable, because the kernel π_n is measurable and the space of all Dirac measures on $\mathbf{A}$ is measurable. (It is closed in the weak topology, if we introduce the proper separable and metrizable topology in $\mathbf{A}$ generating the σ-algebra $\mathcal{B}(\mathbf{A})$.) Therefore, function $\varphi_n(h_{n-1}, \xi_n, u)$ from Theorem 1 can be extended to $\mathbf{H}_{n-1} \times \mathbf{\Xi} \times \mathbb{R}_+$ in a measurable way. The obtained ξ-strategy S' is the desired one because, for each $n \in \mathbb{N}$ for almost all $u \in \mathbb{R}_+$, $\pi_n(\cdot|H_{n-1}, \Xi_n, u) = \delta_{\varphi(H_{n-1},\Xi_n,u)}(\cdot)$ P^S_γ-a.s.

Conversely, suppose the control strategy S is indistinguishable from a ξ-strategy S' defined by the functions $\varphi_n(\cdot)$ (7). For the ξ-strategy S', one can take the common probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ for all $n \in \mathbb{N}$, $h_{n-1} \in \mathbf{H}_{n-1}$, $\xi_n \in \mathbf{\Xi}$, namely, the trivial space with $\tilde{\Omega} = \{\tilde{\omega}\}$ being a singleton. Definition 3 holds true for the process $A(u, \tilde{\omega}) = \varphi_n(h_{n-1}, \xi_n, u)$. The original strategy S is realizable due to Remark 1. □

According to the proof of Corollary 1(b), we see that the process

$$\varphi(t, \omega) = \sum_{n=1}^{\infty} I\{T_{n-1}(\omega) < t \le T_n(\omega)\}\varphi_n(H_{n-1}(\omega), \Xi_n(\omega), t - T_{n-1}(\omega)) \tag{17}$$

equivalently represents the ξ-strategy S, defined by the functions $\varphi_n(\cdot)$, in the sense that for all $n \in \mathbb{N}$, for P^S_γ-almost all (H_{n-1}, Ξ_n) with $T_{n-1} < \infty$, $X_{n-1} \in \mathbf{X}$, the following assertions are valid:

- for any measurable cost function $c(\cdot)$, the cost rate on $(T_{n-1}, T_n]$ is $c(X_{n-1}, \varphi(u))$ and
- for any transition rate $\hat{q}(\cdot)$, the joint distribution of (Θ_n, X_n) is defined by the transition density $\hat{q}(dy|X_{n-1}, \varphi(T_{n-1} + \theta))$.

Below, we fix an arbitrary element $\hat{a} \in \mathbf{A}$ and put $\varphi(t, \omega) \equiv \hat{a}$ for $t \ge T_\infty(\omega)$.

This motivates the following definition. Up to the end of the current section, the space $\mathbf{\Xi}$ is assumed to be arbitrarily fixed.

Definition 4 *Suppose $(\tilde{\Omega}, \tilde{\mathcal{F}})$ is a standard Borel space and put $\hat{\mathbf{\Omega}} = \Omega \times \tilde{\Omega}$, $\hat{\mathcal{F}} = \mathcal{F} \otimes \tilde{\mathcal{F}}$. Let $\hat{P}$ be a probability measure on $\hat{\mathbf{\Omega}}$. A measurable random process $\mathbf{A}(t, \hat{\omega}) = \mathbf{A}(t, (\omega, \tilde{\omega}))$ on $\mathbb{R}_+ \times \hat{\mathbf{\Omega}}$ with values in $\mathbf{A}$ is said to equivalently represent a strategy S if the following assertions are valid.*

- *Leb $\times \hat{P}$-a.s., the process $\mathbf{A}(\cdot)$ has the form*

  $$\mathbf{A}(t, \hat{\omega}) = \mathbf{A}(t, (\omega, \tilde{\omega})) = \sum_{n=1}^{\infty} I\{T_{n-1}(\omega) < t \le T_n(\omega)\}A_n(H_{n-1}(\omega), \Xi_n(\omega), t - T_{n-1}(\omega), \tilde{\omega})$$

  for $t < T_\infty(\omega)$, where $A_n(\cdot)$ is a measurable map from $\mathbf{H}_{n-1} \times \mathbf{\Xi} \times \mathbb{R}_+ \times \tilde{\Omega}$ to $\mathbf{A}$ for all $n \in \mathbb{N}$. For $t \ge T_\infty(\omega)$, $\mathbf{A}(t, (\omega, \tilde{\omega})) \equiv \hat{a}$.
- *The marginal measure $\hat{P}(\Gamma \times \tilde{\Omega})$ coincides with the measure P^S_γ on Ω.*
- *For any non-negative measurable function $c(\cdot)$ on $\mathbf{X} \times \mathbf{A}$, for all $n \in \mathbb{N}$, the actual cost rate*

  $$\int_{\mathbf{A}} I\{X_{n-1} \in \mathbf{X}\}c(X_{n-1}, a)\pi_n(da|H_{n-1}, \Xi_n, u)$$

  on $(T_{n-1}, T_n] \neq \emptyset$ coincides with

  $$\int_{\tilde{\Omega}} I\{X_{n-1} \in \mathbf{X}\}c(X_{n-1}, \mathbf{A}(T_{n-1} + u, (\omega, \tilde{\omega}))\hat{P}(d\tilde{\omega}|\omega)$$

  P^S_γ-a.s. for almost all $u \in \mathbb{R}_+$.

- *For any transition rate $\hat{q}(\cdot)$, for each $n \in \mathbb{N}$, the actual joint distribution (10) of (Θ_n, X_n) given by*

$$I\{X_{n-1} \in \mathbf{X}\}\left\{\int_{\Gamma_\mathbb{R}\cap\mathbb{R}_+} G_n^{\xi,\hat{q}}(\Gamma_\mathbf{X}, H_{n-1}, \Xi_n, \theta)d\theta + I\{\infty \in \Gamma_\mathbb{R}\}I\{\Delta \in \Gamma_\mathbf{X}\}e^{-\Lambda_n^{\hat{q}}(H_{n-1},\Xi_n,\infty)}\right\}$$

coincides with

$$\int_{\tilde{\Omega}} I\{X_{n-1} \in \mathbf{X}\}\left\{\int_{\Gamma_\mathbb{R}\cap\mathbb{R}_+} \hat{q}(\Gamma_\mathbf{X} \setminus \{\Delta, X_{n-1}\}|X_{n-1}, \mathbf{A}(T_{n-1}+\theta, (\omega,\tilde{\omega})))\right.$$
$$\times e^{-\int_{(0,\theta]}\hat{q}_{X_{n-1}}(\mathbf{A}(T_{n-1}+u,(\omega,\tilde{\omega})))du}d\theta$$
$$\left.+I\{\infty \in \Gamma_\mathbb{R}\}I\{\Delta \in \Gamma_\mathbf{X}\}e^{-\int_{(0,\infty)}\hat{q}_{X_{n-1}}(\mathbf{A}(T_{n-1}+u,(\omega,\tilde{\omega})))du}\right\}\hat{P}(d\tilde{\omega}|\omega)$$

P_γ^S*-a.s. Here* $\Gamma_\mathbb{R} \in \mathcal{B}(\bar{\mathbb{R}}_+)$, $\Gamma_\mathbf{X} \in \mathcal{B}(\mathbf{X}_\Delta)$.

Just for brevity, when we say that a process $\mathbf{A}(\cdot)$ equivalently represents a strategy S, we assume also that the space $(\tilde{\Omega}, \tilde{\mathcal{F}})$ and the probability measure $\hat{P}$ on $\mathbf{\Omega} = \Omega \times \tilde{\Omega}$ are fixed.

Remark 2 *According to the discussion of formula (17), if S is a ξ-strategy then, after we build the process $\varphi(\cdot)$ and put $\tilde{\Omega} = \{\tilde{\omega}\}$, $\hat{P}(\Gamma \times \{\tilde{\omega}\}) = P_\gamma^S(\Gamma)$ for $\Gamma \in \mathcal{F}$, the process $\mathbf{A}(t, (\omega, \tilde{\omega})) = \varphi(t, \omega)$ equivalently represents the strategy S.*

Remark 3 *If a process $\mathbf{A}(\cdot)$ equivalently represents a strategy S, then it also equivalently represents any strategy indistinguishable from S.*

Theorem 2 *A control strategy S is realizable if and only if there exists a process $\mathbf{A}(\cdot)$ which equivalently represents S.*

Proof. Suppose a control strategy S is realizable. By Corollary 1(b) and Remark 2, there exists a process $\mathbf{A}(\cdot)$ which equivalently represents the ξ-strategy S' indistinguishable from S. That process equivalently represents also the initial strategy S by Remark 3.

Suppose a process $\mathbf{A}(\cdot)$ equivalently represents a strategy S and show that S is realizable. The only difficulty is to construct the measure $\tilde{P}$ for a fixed $(h_{n-1}, \xi_n) \in \mathbf{H}_{n-1} \times \mathbf{\Xi}$ with $t_{n-1} < \infty$ ($n \in \mathbb{N}$). Let $\mathcal{F}_n = \sigma(H_{n-1}, \Xi_n)$ and consider the restriction of the measure $\hat{P}$ on $(\mathcal{F}_n \otimes \tilde{\mathcal{F}})$. After we disintegrate it, we obtain the $\mathcal{F}_n$-measurable stochastic kernel on $\tilde{\Omega}$ given $\omega \in \Omega$ which has the form $\tilde{P}(d\tilde{\omega}|H_{n-1}(\omega), \Xi_n(\omega))$. Now, for P_γ^S-almost all (H_{n-1}, Ξ_n) with $T_{n-1} < \infty$, the requirements (a) and (b) of Definition 3 hold true for the process $A_n(H_{n-1}, \Xi_n, u, \tilde{\omega})$ and measure $\tilde{P}(d\tilde{\omega}|H_{n-1}, \Xi_n)$. To check (a), it is sufficient to put $c(x, a) = I\{a \in \Gamma_\mathbf{A}\}$; for each h_{n-1}, ξ_n, the measure $\tilde{P}(d\tilde{\omega}|h_{n-1}, \xi_n)$ can be completed if needed. □

Let us look more attentively at the processes $\mathbf{A}(\cdot)$ which equivalently represent this or that control strategy under a fixed space $\mathbf{\Xi}$. Collection of all such processes is denoted as $\aleph$. The space $\tilde{\Omega}$ may be different for different processes from $\aleph$.

First of all, the process $\varphi(\cdot)$ of the type (17) can be characterized in the following way. On the space Ω, consider random measure

$$\mu(\omega, dt, d(x, \xi)) = \sum_{n=1}^{\infty} \delta_{(T_n(\omega),(X_n(\omega),\Xi_{n+1}(\omega)))}(dt, d(x, \xi))$$

and σ-algebras

$$\begin{aligned}\mathcal{F}_0 &= \sigma(\Xi_0, X_0, \Xi_1);\\ \mathcal{F}_t &= \mathcal{F}_0 \vee \sigma(\mu((0, s] \times B) :\ s \le t,\ B \in \mathcal{B}(\mathbf{X} \times \mathbf{\Xi})).\end{aligned}$$

The associated σ-algebra on $\Omega \times \mathbb{R}_+^0$ is defined as

$$\sigma\left(\Gamma \times \{0\}\ (\Gamma \in \mathcal{F}_0), \quad \Gamma \times (s, \infty)\ (\Gamma \in \bigvee_{t<s} \mathcal{F}_t,\ s > 0)\right).$$

Now a random process $\varphi(\cdot)$ on $\Omega \times \mathbb{R}_+$ has the form (17) if and only if it is predictable [16, Lemma(3.3)]. Remember, we agreed to put $\varphi(t,\omega) \equiv \hat{a}$ for $t \geq T_\infty(\omega)$.

Theorem 3 *(a) If the process $\mathbf{A}(\cdot)$ equivalently represents a strategy S (i.e., $\mathbf{A}(\cdot) \in \aleph$), then there exists a predictable process $\varphi(\cdot)$ on $\Omega \times \mathbb{R}_+$ such that*

$$Leb \times \hat{P}\text{-a.s.} \quad \mathbf{A}(t,(\omega,\tilde{\omega})) = \varphi(t,\omega),$$

and the process $\mathbf{A}'(t,(\omega,\tilde{\omega}')) = \varphi(t,\omega)$ also equivalently represents the strategy S. Here $\tilde{\Omega}' = \{\tilde{\omega}'\}$ is a singleton and $\hat{P}(\Gamma \times \{\tilde{\omega}'\}) = P^S_\gamma(\Gamma)$ for $\Gamma \in \mathcal{F}$.

(b) For each predictable process $\varphi(\cdot)$, there is a probability measure $\hat{P}$ on $\hat{\mathbf{\Omega}} = \Omega \times \{\tilde{\omega}\}$ such that the process $\mathbf{A}(t,(\omega,\tilde{\omega})) = \varphi(t,\omega)$ equivalently represents some strategy S (i.e., $\mathbf{A}(\cdot) \in \aleph$).

Proof. (a) According to Theorem 2, Corollary 1(b) and Remark 3, the $\mathbf{A}(\cdot)$ process equivalently represents the ξ-strategy S' indistinguishable from S and defined by functions $\varphi_n(\cdot)$ (7), which give rise to the process (17). According to the second part of the proof of Theorem 2 applied to the strategy S', for each $n \in \mathbb{N}$, with $P^{S'}_\gamma$-probability one, the requirements (a) and (b) of Definition 3 hold true for (H_{n-1}, Ξ_n) with $T_{n-1} < \infty$, for the process $A_n(H_{n-1}, \Xi_n, u, \tilde{\omega})$ and measure $\tilde{P}(d\tilde{\omega}|H_{n-1}, \Xi_n)$. By Corollary 1 (a), for almost all $u \in \mathbb{R}_+$

$$\tilde{P}(A_n(H_{n-1}, \Xi_n, u, \tilde{\omega}) = \varphi_n(H_{n-1}, \Xi_n, u)|H_{n-1}, \Xi_n) = 1. \tag{18}$$

Note that we deal with the ξ-strategy S', so that $\pi_n(\cdot|h_{n-1}, \xi_n, u) = \delta_{\varphi_n(h_{n-1},\xi_n,u)}(\cdot)$ for all $n \in \mathbb{N}$, $(h_{n-1}, \xi_n, u) \in \mathbf{H}_{n-1} \times \mathbf{\Xi} \times \mathbb{R}_+$. Equality (18) holds $P^{S'}_\gamma$-a.s. (and also P^S_γ-a.s.). Hence $Leb \times \hat{P}$-a.s. $\mathbf{A}(t,(\omega,\tilde{\omega})) = \varphi(t,\omega)$.

According to Remark 2, the process $\mathbf{A}'(\cdot)$ equivalently represents the strategy S' as well as the initial strategy S indistinguishable form S'. (See Remark 3.) Remember, $P^S_\gamma = P^{S'}_\gamma$.

(b) The process $\varphi(\cdot)$ has the form (17). Fix an arbitrary probability $p_0(\cdot)$ on $\mathbf{\Xi}$ and arbitrary stochastic kernels $p_n(\cdot)$ on $\mathbf{\Xi}$ given $\mathbf{H}_{n-1}$, $n \in \mathbb{N}$, and consider the corresponding ξ-strategy S defined by the maps $\varphi_n(\cdot)$. It remains to put $\hat{P}(\Gamma \times \{\tilde{\omega}\}) = P^S_\gamma(\Gamma)$ for $\Gamma \in \mathcal{F}$ and refer to Remark 2. □

Definition 5 *We say that two processes $\mathbf{A}_1(\cdot), \mathbf{A}_2(\cdot)$ from $\aleph$ belong to the same class if there exists one ξ-strategy equivalently represented by both $\mathbf{A}_1(\cdot)$ and $\mathbf{A}_2(\cdot)$.*

Theorem 4 *The classes form a partition of $\aleph$, that is,*

- *each process from $\aleph$ belongs to some class;*
- *two classes either coincide or do not overlap.*

Proof. If $\mathbf{A}(\cdot) \in \aleph$ then there is a ξ-strategy equivalently representable by $\mathbf{A}(\cdot)$ due to Theorem 2, Corollary 1(b) and Remark 3.

Suppose the process $\mathbf{A}(\cdot)$ equivalently represents two ξ-strategies S^1 and S^2 defined by the maps $\varphi^1_n(\cdot)$ and $\varphi^2_n(\cdot)$ (7). We will show that the strategies S^1 and S^2 are indistinguishable. By Definition 4, $P^{S^1}_\gamma = P^{S^2}_\gamma$, and assertion (a) of Definition 2 follows, because the components $p_0(\cdot)$, $p_n(\cdot)$ ($n \in \mathbb{N}$) can be constructed starting from the strategical measure $P^{S^1}_\gamma = P^{S^2}_\gamma$. Let a non-negative measurable function $c(\cdot)$ on $\mathbf{A}$ be such that $c(a_1) \neq c(a_2)$ if $a_1 \neq a_2$. Such a function exists because the standard Borel space $\mathbf{A}$ is isomorphic to the segment $[0,1]$ or its countable subset [3, Cor.7.16.1]. Now again by Definition 4, for each $n = 1, 2, \ldots$, on the set $(T_{n-1}, T_n] \neq \emptyset$, $\varphi^1_n(H_{n-1}, \Xi_n, u) = \varphi^2_n(H_{n-1}, \Xi_n, u)$ $P^{S^1}_\gamma$-a.s. for almost all $u \in \mathbb{R}_+$. Hence the strategies S^1 and S^2 are indistinguishable and the class associated with S^1 coincides with the class associated with S^2. (See Remark 3.) □

Now it is clear that there is 1-1 correspondence between the introduced equivalence classes of the processes and the equivalence classes of indistinguishable realizable strategies. Indeed, in each such class of strategies, say, Π', there is at least one ξ-strategy by Corollary 1(b). The associated class of processes contains those and only those processes which equivalently represent all the strategies from Π': see Remark 3 and the proof of Theorem 4.

By Theorem 3, in each equivalence class of processes $\mathbf{A}(\cdot)$ from $\aleph$, there is at least one canonical process, i.e., such a process that $\tilde{\Omega} = \{\tilde{\omega}\}$ is a singleton and the process $\mathbf{A}(t, (\omega, \tilde{\omega}))$ on $\Omega \times \mathbb{R}_+$ is predictable. This process equivalently represents every strategy from the corresponding class of indistinguishable realizable strategies. (See Remark 3.) Therefore, canonical processes are sufficient if we are looking for a set of processes which equivalently represent realizable strategies.

If $\Xi = \{\tilde{\xi}\}$ is a singleton, then the argument $\tilde{\xi}$ can be omitted everywhere, and the model transforms to the classical continuous-time Markov decision process [10, 11, 12, 13, 18, 20, 24]. In this case, by Corollary 1(b), a strategy is realizable if and only if it is indistinguishable from a randomized strategy which is defined by functions $\varphi_n(h_{n-1}, u)$ and is actually deterministic.

4 Sufficient classes of realizable strategies

As was proved in [22, Th.2], for any strategy S, there is a relaxed Markov strategy S^π (i.e., all the kernels $\pi_n(\cdot|h_{n-1}, \xi_n, u)$ look like $\pi^M_n(\cdot|x_{n-1}, u)$) such that $\{\eta^S_n\}_{n=1}^\infty = \{\eta^{S^\pi}_n\}_{n=1}^\infty$. One can find the explicit expression for that S^π strategy in [22, p.3520]. Therefore, relaxed strategies form a sufficient class for problem (16). But as was shown, they are usually non-realizable.

A simplest realizable strategy, below called Markov standard ξ-strategy, is defined by $\Xi = \mathbf{A}$, $p_n(da_n|h_{n-1}) = p^M_n(da_n|x_{n-1})$, $\varphi(h_{n-1}, a_n, u) = a_n$: the decision maker, at every jump epoch, chooses the randomized action a_n depending on the current state and the jump number, and that action remains constant up to the next one jump. If the kernels $p^M_n(\cdot)$ do not depend on $n \in \mathbb{N}$, then the standard ξ-strategy is called stationary. According to [22, Th.1], if $q_x(a) > 0$ for all $x \in \mathbf{X}$, $a \in \mathbf{A}$, then, for any $S \in \Pi$, there is a Markov standard ξ-strategy S^ξ such that $\eta^{S^\xi}_n \geq \eta^S_n$ for all $n \in \mathbb{N}$. One can find the explicit expression for that S^ξ strategy in [22, p.3519]. Hence, in this case, Markov standard ξ-strategies form a sufficient class for problem (16) with negative cost rates c^j. They are sufficient for arbitrary cost rates c^j if $q_x(a) \geq \varepsilon$ for some $\varepsilon > 0$ [22, Th.1]. Example 2 in [22] shows that, in case $q_x(a) > 0$ and $c^j_n(x, a) > 0$, the infimum in (16) over all strategies can be strictly smaller than the infimum over Markov standard ξ-strategies.

Note that the discounted cost model [8, 20] means, there is a positive rate α of the transitions to a cemetery from each state $x \in \mathbf{X}$. Thus, $q_x(a) \geq \alpha > 0$, and the previous reasoning applies. In this special case, the sufficiency of Markov standard ξ-strategies was established in [8], although using different notations and constructions.

It is easy to understand, why the Markov standard ξ-strategies and even their history-dependent modifications are not sufficient in case $c^j > 0$, if we consider the following trivial model: $\mathbf{X} = \{1\}$, $\mathbf{A} = (0, 1]$, $q_x(a) = 0$, $c^0(x, a) = a$, $J = \emptyset$ (unconstrained case), $\gamma(\{1\}) = 1$. For every Markov standard ξ-strategy S^ξ, when the action is fixed until the next one jump, i.e., fixed forever, $W^0(S^\xi) = \infty$. But, if, e.g., one applies action $a_1 = \frac{\delta}{2}$ on the interval $(0, 1]$, action $a_2 = \frac{\delta}{4}$ on the interval $(1, 2]$, and so on, then, for such strategy S, which is again a special case of randomized, hence realizable, we have $W^0(S) = \delta$, and $\delta > 0$ can be arbitrary. Similar examples in the theory of discrete-time Markov decision processes are well known [21, §2.2.11]. Thus, it is reasonable to consider ξ-strategies similar to the Markov standard ones, but where it is allowed to change actions at some time moments between the jumps. To be more specific, after the $(n-1)$-th jump of the original process $X(\cdot)$ at the time moment T_{n-1}, we allow to choose different actions $A^n_1, A^n_2, \ldots$ on the intervals $(T_{n-1}, T_{n-1}+T^n_1]$, $(T_{n-1}+T^n_1, T_{n-1}+T^n_1+T^n_2], \ldots$ correspondingly. When considering the process $X(\cdot)$ only at those moments T_{n-1}, $T_{n-1} + T^n_1$, $T_{n-1} + T^n_2$, …, one can talk about the standard discrete-time Markov decision process with the corresponding actions $A^n_1, A^n_2, \ldots$. Such construction is presented in Section 5. The described idea, when the mentioned time moments $T^n_1, T^n_2, \ldots$ form an independent of the past Poisson process, leads to the following definition.

Definition 6 *A Poisson-related ξ-strategy $S^P = \{\Xi, p_0, \langle p_n, \pi_n \rangle, n = 1, 2, \ldots\}$ is defined by $\{\Xi = (\mathbb{R} \times \mathbf{A})^\infty = \{(\alpha_1, \tau_1, \alpha_2, \tau_2, \ldots)\}, \varepsilon > 0, \tilde{p}_{n,k}(da|x_{n-1}), n \in \mathbb{N}, k \in \mathbb{N}\}$, where $\tilde{p}_{n,k}$ are stochastic kernels on $\mathbf{A}$ given $\mathbf{X}$, in the following way. For $n = 1, 2, \ldots$ the distribution p_n of $\Xi_n = (A^n_1, T^n_1, A^n_2, \ldots)$ given $\mathbf{H}_{n-1}$ is as follows:*

- *for all $k \geq 1$, $p_n(T^n_k \leq t|h_{n-1}) = 1 - e^{-\varepsilon t}$; exponential random variables T^n_k are mutually independent and independent of $\mathcal{F}_{T_{n-1}} = \mathcal{F}(H_{n-1})$;*
- *for all $k \geq 1$, $p_n(A^n_k \in \Gamma_{\mathbf{A}}|h_{n-1}) = \tilde{p}_{n,k}(\Gamma_{\mathbf{A}}|x_{n-1})$, and, given x_{n-1}, the random elements A^n_k are mutually independent and independent of $(\Xi_0, X_1, \Xi_1, \Theta_1, X_1, \ldots, \Xi_{n-1}, \Theta_{n-1})$ and of $T^n_1, T^n_2, \ldots$.*

Finally,

$$\pi_n(da|h_{n-1}, \xi_n, u) = \delta_{\varphi_n(h_{n-1}, \xi_n, u)}(da),$$

where

$$\varphi_n(h_{n-1}, \xi_n, u) = \sum_{k=1}^{\infty} I\{\tau^n_1 + \ldots + \tau^n_{k-1} < u \leq \tau^n_1 + \ldots + \tau^n_k\}\alpha^n_k.$$

As usual, $(\alpha^n_1, \tau^n_1, \alpha^n_2, \ldots) = \xi_n \in \mathbf{\Xi}$ denotes the realization of the random element $(A^n_1, T^n_1, A^n_2, \ldots) = \Xi_n$. The ξ_0 component plays no role and can be omitted. Note, the function φ_n does not depend on h_{n-1}.

Such a strategy means that, after any jump of the controlled process $X(\cdot)$, the decision maker simulates a Poisson process and applies different randomized actions during the different sojourn times of that Poisson process. One can say that a Poisson-related ξ strategy is a randomized switching strategy [8] at random time moments. According to [22, Th.5], for any control strategy $S \in \Pi$, there is a Poisson-related ξ-strategy S^P such that $\{\eta^S_n\}^\infty_{n=1} = \{\eta^{S^P}_n\}^\infty_{n=1}$. One can find the explicit expression of that S^P strategy in [23, p.199-200] and also in [22, p.3527]. Now it is clear that the class of Poisson-related ξ-strategies is sufficient for problem (16). The value of $\varepsilon > 0$ can be chosen arbitrarily (but the kernels $\tilde{p}_{n,k}$ depend on ε). Remember, like every randomized strategy, Poisson-related ξ-strategy is realizable. Note also that if $\varepsilon = 0$, then the notion of the Poisson-related strategy transforms to the Markov standard ξ-strategy, because $T^n_k \equiv \infty$. Similarly, if all the kernels $\tilde{p}_{n,k}(\cdot)$ are identical and degenerate, then we actually deal with a stationary standard ξ-strategy. As was mentioned, π-strategies form another sufficient class of strategies because, for any control strategy $S \in \Pi$, there is a π-strategy S^π such that $\{\eta^S_n\}^\infty_{n=1} = \{\eta^{S^\pi}_n\}^\infty_{n=1}$ [22, Th.2]. But π-strategies are usually not realizable.

5 Continuous and discrete-time models with the total expected cost

In this section, we accept that the cost rates c^j do not depend on n, the jump number.

Since Poisson-related ξ-strategies are sufficient for problem (16), let us fix such a strategy S^P and, for each $\omega \in \Omega$, consider the sequence of realized essential time-moments, states and actions:

$$\begin{array}{l}(t_0 = 0, y_0 = x_0, b_1 = \alpha^1_1), (t_0 + \tau^1_1, y_1 = x_0, b_2 = \alpha^1_2), \ldots, (t_0 + \sum_{i=1}^{k_0-1} \tau^1_i, y_{k_0-1} = x_0, b_{k_0} = \alpha^1_{k_0}),\\ (t_1, y_{k_0} = x_1, b_{k_0+1} = \alpha^2_1), (t_1 + \tau^2_1, y_{k_0+1} = x_1, b_{k_0+2} = \alpha^2_2), \ldots,\\ (t_1 + \sum_{i=1}^{k_1-1} \tau^2_i, y_{k_0+k_1-1} = x_1, b_{k_0+k_1} = \alpha^2_{k_1}),\\ (t_2, y_{k_0+k_1} = x_2, b_{k_0+k_1+1} = \alpha^3_1), \ldots\end{array} \tag{19}$$

Here $k_0, k_1, \ldots \geq 1$ and $\tau^n_0 \stackrel{\triangle}{=} 0$. The corresponding time moments are naturally ordered:

$$t_1 \in (t_0 + \sum_{i=1}^{k_0-1} \tau^1_i, t_0 + \sum_{i=1}^{k_0} \tau^1_i),\ t_2 \in (t_1 + \sum_{i=1}^{k_1-1} \tau^2_i, t_1 + \sum_{i=1}^{k_1} \tau^2_i),\ \ldots$$

Recall that $t_n = \sum_{i=1}^n \theta_i$ denote the realized jump moments, and τ^n_k are the realized exponential random variables, the components of ξ_n. If t_n is the last jump moment (the state x_n is absorbing) then $k_n = \infty$: the tail of the introduced sequence is

$$\begin{array}{l}(t_n, y_{k_0+\ldots+k_{n-1}} = x_n, b_{k_0+\ldots+k_{n-1}+1} = \alpha^{n+1}_1),\\ (t_n + \tau^{n+1}_1, y_{k_0+\ldots+k_{n-1}+1} = x_n, b_{k_0+\ldots+k_{n-1}+2} = \alpha^{n+1}_2),\\ \ldots\ \ldots\end{array}$$

Using these sequences, the objective (14) can be represented as

$$\begin{aligned} W^j(S^P) &= \sum_{m=0}^{\infty} E_\gamma^{S^P}[c^{j+}(Y_m, B_{m+1})/(q_{Y_m}(B_{m+1})+\varepsilon)] \\ &\quad + \sum_{m=0}^{\infty} E_\gamma^{S^P}[c^{j-}(Y_m, B_{m+1})/(q_{Y_m}(B_{m+1})+\varepsilon)]. \end{aligned} \tag{20}$$

To show this, note first that, for fixed $n \in \mathbb{N}$ and for $T_{n-1} < \infty$,

$$\begin{aligned} & E_\gamma^{S^p}\left[\int_{(T_{n-1},T_n]\cap\mathbb{R}_+}\int_{\mathbf{A}} \pi_n(da|H_{n-1}, \Xi_n, t-T_{n-1})c^{j+}(X_{n-1}, a)dt\right] \qquad (21) \\ = \; & E_\gamma^{S^p}\left[\sum_{m=K_0+K_1+\ldots+K_{n-2}}^{K_0+K_1+\ldots+K_{n-1}-1} c^{j+}(Y_m, B_{m+1})\hat{T}_{m-(K_0+K_1+\ldots+K_{n-2})+1}\right], \end{aligned}$$

where $K_{-1} = 0$; for $l = 1, 2, \ldots, K_{n-1}$

$$\begin{aligned} \hat{T}_l &= \begin{cases} T_l^n, & \text{if } 1 \le j < K_{n-1}; \\ \Theta_n - (T_1^n + T_2^n + \ldots + T_{K_{n-1}-1}^n), & \text{if } j = K_{n-1} \end{cases} \\ &= \min\{T_l^n,\ \Theta_n - (T_1^n + T_2^n + \ldots + T_{l-1}^n)\} \end{aligned}$$

and $T_0^n = 0$. As usual, the capital letters $K_{n-1}, Y_m, B_{m+1}, T_l^n, \Theta_n$ denote the random elements whose realizations were denoted as $k_{n-1}, y_m, b_{m+1}, \tau_l^n, \theta_n$. For $l = 1, 2, \ldots$ consider

$$E_\gamma^{S^p}\left[c^{j+}(Y_{l+K_0+K_1+\ldots+K_{n-2}-1}, B_{l+K_0+K_1+\ldots+K_{n-2}})\hat{T}_l I\{l \le K_{n-1}\}|\mathcal{G}_l\right],$$

where $\mathcal{G}_l = \mathcal{F}(H_{n-1}) \vee \mathcal{F}(A_1^n, T_1^n, \ldots, A_l^n)$. Note that $Y_{l+K_0+K_1+\ldots+K_{n-2}-1} = X_{n-1}$, $B_{l+K_0+K_1+\ldots+K_{n-2}} = A_l^n$ and $I\{l \le K_{n-1}\} = I\{\Theta_n \ge T_1^n + T_2^n + \ldots + T_{l-1}^n\}$. For fixed values of $X_{n-1} = x_{n-1}$, $A_1^n = \alpha_1^n$, $T_1^n = \tau_1^n$, …, $A_l^n = \alpha_l^n$, under the condition $\Theta_n \ge \tau_1^n + \tau_2^n + \ldots + \tau_{l-1}^n$, the random variables $\Delta\Theta = \Theta_n - (\tau_1^n + \tau_2^n + \ldots + \tau_{l-1}^n)$ and T_l^n have densities $q_{x_{n-1}}(\alpha_l^n)e^{-q_{x_{n-1}}(\alpha_l^n)z}$ (when $z \le T_l^n$ and assuming $q_{x_{n-1}}(\alpha_l^n) > 0$) and $\varepsilon e^{-\varepsilon z}$ correspondingly. Therefore (also in the case $q_{x_{n-1}}(\alpha_l^n) = 0$) the conditional expectation $E_\gamma^{S^P}\left[\hat{T}_l I\{l \le K_{n-1}\}|\mathcal{G}_l\right]$ equals the expectation of the minimum of two independent exponential random variables:

$$\begin{aligned} & E_\gamma^{S^P}\left[c^{j+}(Y_{l+K_0+K_1+\ldots+K_{n-2}-1}, B_{l+K_0+K_1+\ldots+K_{n-2}})\hat{T}_l I\{l \le K_{n-1}\}|\mathcal{G}_l\right] \\ = \; & c^{j+}(X_{n-1}, A_l^n)\frac{1}{q_{X_{n-1}}(A_l^n)+\varepsilon}, \end{aligned}$$

and the expression (21) equals

$$E_\gamma^{S^p}\left[\sum_{m=K_0+K_1+\ldots+K_{n-2}}^{K_0+K_1+\ldots+K_{n-1}-1} c^{j+}(Y_m, B_{m+1})/(q_{X_{n-1}}(B_{m+1})+\varepsilon)\right].$$

The desired formula (20) follows.

The sequence

$${}^{\mathcal{M}}\omega = (y_0, b_1, y_1, b_2, \ldots, y_m, b_{m+1}, y_{m+1}, \ldots)$$

is in fact a trajectory of a discrete-time Markov decision process. Indeed, for any $y_m \in \mathbf{X}$, $b_{m+1} \in \mathbf{A}$, the value of y_{m+1} is the realization of the random element with distribution

$$Q(\Gamma|y_m, b_{m+1}) = \frac{q(\Gamma \setminus \{y_m\}|y_m, b_{m+1}) + \varepsilon I\{\Gamma \ni y_m\}}{q_{y_m}(b_{m+1})+\varepsilon}, \quad \Gamma \in \mathcal{B}(\mathbf{X}), \quad m \in \mathbb{N}.$$

This discrete-time Markov decision process is denoted as $\mathcal{M}$, and the histories, strategical measures etc, relevant to $\mathcal{M}$, are usually equipped with the left upper index $\mathcal{M}$.

For each $m \in \mathbb{N} \cup \{0\}$, having in hand the history ${}^{\mathcal{M}}h_m = (y_0, b_1, y_1, \ldots, b_m, y_m)$ in $\mathcal{M}$, one can recalculate the values of $n \geq 1$ and $k \geq 1$ such that

$$\begin{aligned}
&(t_0 = 0, y_0 = x_0, b_1 = \alpha_1^1), (t_0 + \tau_1^1, y_1 = x_0, b_2 = \alpha_2^1), \ldots, (t_0 + \tau_{k_0-1}^1, y_{k_0-1} = x_0, b_{k_0} = \alpha_{k_0}^1),\\
&(t_1, y_{k_0} = x_1, b_{k_0+1} = \alpha_1^2), (t_1 + \tau_1^2, y_{k_0+1} = x_1, b_{k_0+2} = \alpha_2^2), \ldots,\\
&(t_1 + \tau_{k_1-1}^2, y_{k_0+k_1-1} = x_1, b_{k_0+k_1} = \alpha_{k_1}^2), (t_2, y_{k_0+k_1} = x_2, b_{k_0+k_1+1} = \alpha_1^3), \ldots,\\
&(t_{n-1} + \tau_{k-2}^n, y_{m-1} = x_{n-1}, b_m = \alpha_{k-1}^n), (t_{n-1} + \tau_{k-1}^n, y_m = x_{n-1}) \text{ (if } y_m = y_{m-1}\text{; then } k \geq 2),
\end{aligned}$$

or

$$\begin{aligned}
&(t_0 = 0, y_0 = x_0, b_1 = \alpha_1^1), (t_0 + \tau_1^1, y_1 = x_0, b_2 = \alpha_2^1), \ldots, (t_0 + \tau_{k_0-1}^1, y_{k_0-1} = x_0, b_{k_0} = \alpha_{k_0}^1),\\
&(t_1, y_{k_0} = x_1, b_{k_0+1} = \alpha_1^2), (t_1 + \tau_1^2, y_{k_0+1} = x_1, b_{k_0+2} = \alpha_2^2), \ldots,\\
&(t_1 + \tau_{k_1-1}^2, y_{k_0+k_1-1} = x_1, b_{k_0+k_1} = \alpha_{k_1}^2), (t_2, y_{k_0+k_1} = x_2, b_{k_0+k_1+1} = \alpha_1^3), \ldots,\\
&(t_{n-1} + \tau_{k_{n-1}-1}^n, y_{m-1} = x_{n-1}, b_m = \alpha_{k_{n-1}}^n), (t_n, y_m = x_n) \text{ (if } y_m \neq y_{m-1}\text{; then } k = 1).
\end{aligned}$$

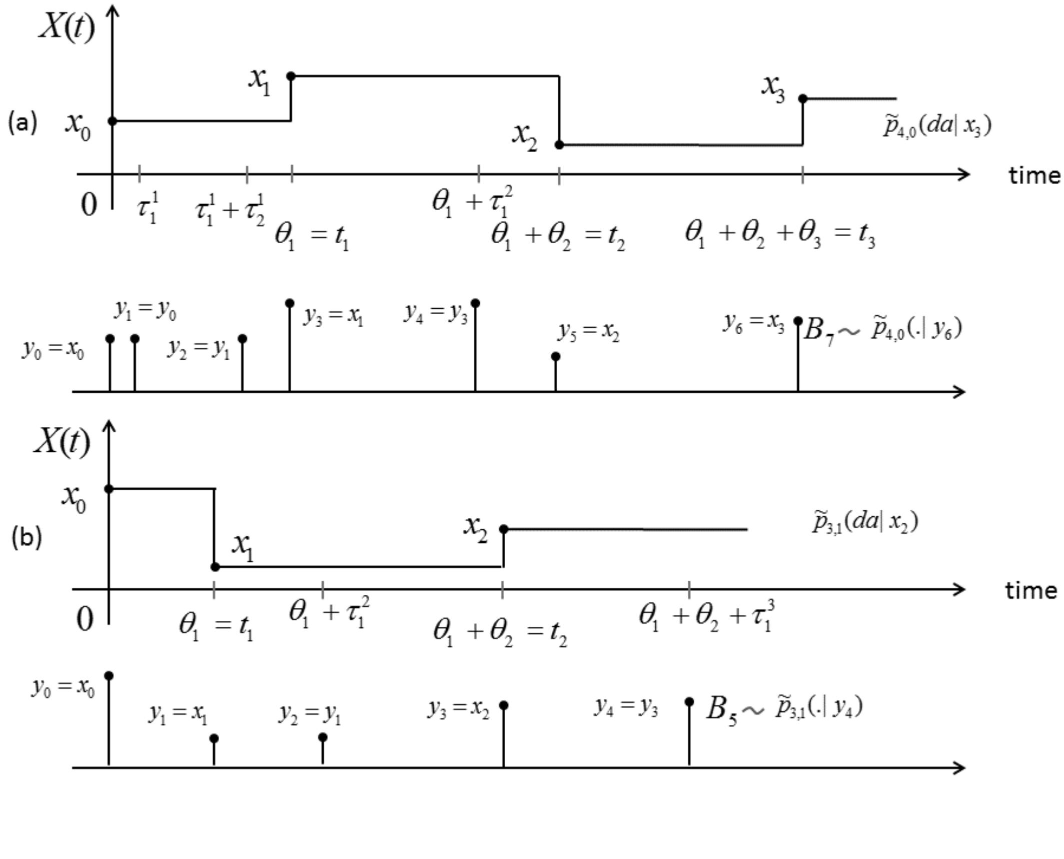

Figure 1: Two scenarios illustrating the construction and the connection between the histories ${}^{\mathcal{M}}h_m$ and the trajectories of the CTMDP:
(a) $m = 6$; ${}^{\mathcal{M}}h_6 = (y_0, b_1, \ldots, y_6)$; $n({}^{\mathcal{M}}h_6) = 4$, $k({}^{\mathcal{M}}h_6) = 0$;
(b) $m = 4$; ${}^{\mathcal{M}}h_4 = (y_0, b_1, \ldots, y_4)$; $n({}^{\mathcal{M}}h_4) = 3$, $k({}^{\mathcal{M}}h_4) = 1$.

To do this, one should simply remember that the value of n (counting the real jumps of the controlled process $X(\cdot)$) increases by 1 every time when the next value of y_m is different from the previous value y_{m-1}. More detailed explanations are in [23].

If we denote as $n({}^{\mathcal{M}}h_m)$ and $k({}^{\mathcal{M}}h_m)$ those values of n and k and apply the control strategy ${}^{\mathcal{M}}S$ in the $\mathcal{M}$ model defined by

$${}^{\mathcal{M}}p_{m+1}(\cdot|{}^{\mathcal{M}}h_m) = \tilde{p}_{n({}^{\mathcal{M}}h_m),k({}^{\mathcal{M}}h_m)}(\cdot|y_m),$$

then the strategical measure $P_\gamma^{{}^{\mathcal{M}}S}$ in $\mathcal{M}$ coincides with the measure $P_\gamma^{S^P}$ on the space of trajectories ${}^{\mathcal{M}}\omega$, so that

$$W^j(S^P) = \sum_{m=0}^{\infty} E_\gamma^{{}^{\mathcal{M}}S}[c^j(Y_m, B_{m+1})/(q_{Y_m}(B_{m+1}) + \varepsilon)], \tag{22}$$

where $E_\gamma^{{}^{\mathcal{M}}S}$ is the mathematical expectation w.r.t. $P_\gamma^{{}^{\mathcal{M}}S}$. These constructions are illustrated on Figure 1.

Conversely, suppose a Markov control strategy ${}^{\mathcal{M}}S$ in $\mathcal{M}$, defined by ${}^{\mathcal{M}}p_{m+1}(\cdot|y_m)$, is fixed ($m \in \mathbb{N} \cup \{0\}$). According to [19, Lemma 2], for an arbitrary strategy in $\mathcal{M}$, there is a Markov strategy such that the objectives (22) coincide for any measurable function c^j, so that we don't loose the generality being restricted to Markov strategies. Firstly, we construct the equivalent, in the sense of (22), strategy in the original continuous-time model, which is a little more general than Poisson-related, namely, history-dependent. That means, the space $\mathbf{\Xi} = (\mathbb{R} \times \mathbf{A})^\infty$ is the same, but the kernels $\tilde{p}_{n,k}$ may depend on h_{n-1}, not only on x_{n-1}. The histories h_n in this case look similarly to the case of a standard Poisson-related strategy. For any such history

$$h_{n-1} = (\xi_0, x_0, \xi_1, \theta_1, x_1, \ldots, \xi_{n-1}, \theta_{n-1}, x_{n-1}), \quad n = 1, 2, \ldots$$

one can build the corresponding history ${}^{\mathcal{M}}h_m$ in $\mathcal{M}$, using the formulae (19), where the last element is ($t_{n-1} = \sum_{i=1}^{n-1} \theta_i, y_m = x_{n-1}$). The corresponding value of m is denoted as $m(h_{n-1}) \in \mathbb{N} \cup \{0\}$. It remains to put

$$\tilde{p}_{n,k}(\cdot|h_{n-1}) = {}^{\mathcal{M}}p_{m(h_{n-1})+k}(\cdot|y_{m(h_{n-1})}), \quad k = 1, 2, \ldots$$

to obtain the history-dependent Poisson-related strategy in the original continuous-time model with the strategical measure coincident with $P_\gamma^{{}^{\mathcal{M}}S}$ on the space of trajectories ${}^{\mathcal{M}}\omega$. For that history-dependent strategy, we take the (standard) Poisson-related strategy S^P leading to the same sequence of the detailed occupation measures $\{\eta_n^{S^P}\}_{n=1}^\infty$ (see Section 4). As a result, we obtain equality (22) valid for any measurable function c^j.

To summarise, for any Poisson-related strategy S^P, there is a Markov strategy ${}^{\mathcal{M}}S$ in $\mathcal{M}$ (and conversely, for any Markov strategy ${}^{\mathcal{M}}S$ in $\mathcal{M}$, there is a Poisson-related strategy S^P) such that equality (22) is valid. Therefore, solving problem (16) is equivalent to solving the discrete-time problem

$$\sum_{m=0}^{\infty} E_\gamma^{{}^{\mathcal{M}}S}[c^0(Y_m, B_{m+1})/(q_{Y_m}(B_{m+1}) + \varepsilon)] \quad \to \inf_{{}^{\mathcal{M}}S \text{ are Markov in } \mathcal{M}} \tag{23}$$
$$\text{subject to } \sum_{m=0}^{\infty} E_\gamma^{{}^{\mathcal{M}}S}[c^j(Y_m, B_{m+1})/(q_{Y_m}(B_{m+1}) + \varepsilon)] \quad \leq \quad d^j, \quad j \in J.$$

Remember also that, if the total transition rate $q_x(a)$ is strictly separated from zero (e.g., one is dealing with the discounted cost model), then one can restrict himself with the Markov standard ξ-strategies which are Poisson-related with $\varepsilon = 0$: see Section 4. All the reasoning in the current section applies in this special case.

Now all the theory, developed for the discrete-time Markov decision processes, can be used for solving problem (16). Without intention to provide an exhaustive survey, let us mention several special cases.

1. If $J = \emptyset$ (the case of the unconstrained optimization) and $c^0(\cdot) \geq 0$, then, under appropriate compactness-continuity conditions, there is an optimal non-randomized stationary strategy

in $\mathcal{M}$ [3, Cor.9.17.2]. That strategy gives rise to the optimal Poisson-related strategy with the degenerate identical stochastic kernels $\tilde{p}_{n,k}$, which in fact is a stationary Markov standard ξ-strategy. The standard dynamic programming approach applies if $J = \emptyset$ [3, 7, 14]. Let us remind that in general, stationary strategies are not sufficient for solving optimal control problems [21, §2.2.11].

2. If $J \neq \emptyset$ is finite then the convex analytic approach to problem (23) leads to the linear program on the space of (total) occupation measures on $\mathbf{X} \times \mathbf{A}$

   $$\mu^{{}^{\mathcal{M}}S}(\Gamma) = \sum_{m=0}^{\infty} P_{\gamma}^{{}^{\mathcal{M}}S}((Y_m, B_{m+1}) \in \Gamma).$$

   One can find the details in [4, 14]. For example, under appropriate conditions, if one succeeds to find a solution to that linear program, then it is possible to construct the so called induced (stationary randomized) strategy in $\mathcal{M}$ solving problem (23) [4, Th.5.2]. That strategy again gives rise to the optimal Poisson-related strategy solving problem (16).

3. For the case of finite or countable state space $\mathbf{X}$, many results about discounted, absorbing, and transient constrained models can be found in [1]. See also [6].

4. The Borel discounted model with constraints was studied in [5, 19]. The weight function technique was demonstrated in [5].

Application of the discrete-time methods to controlled continuous-time models, for instance the policy iteration in the unconstrained case, appeared already in [15] for simple semi-Markov decision processes. For a more general model, reducing to the discrete-time case was described in [27], again in the framework of the unconstrained model and dynamic programming approach. Here, actions in the discrete-time model were $\mathbf{A}$-valued functions. For the multiple discounted objectives, such a reduction was developed in [8, 9] using the concept of occupation measures. Here, actions in the discrete-time model were just the original actions from $\mathbf{A}$.

6 Conclusion

In the recent decades, many authors provided solutions to continuous-time Markov decision processes in the form of (relaxed) π-strategies. As is known, such strategies are usually not realizable on practice. On the other hand, simple realizable randomized strategies, corresponding to the case $\mathbf{\Xi} = \mathbf{A}$ and called Markov standard ξ-strategies in the current paper, are not sufficient for solving optimal control problems. Even their history-dependent modifications are not sufficient. Moreover, such strategies do not fit the definition of a strategy accepted in many articles and books.

In the current paper, following [22], we introduced the most wide class of π-ξ-strategies which makes it possible to study the classical continuous-time Markov decision processes as well as the exponential semi-Markov decision processes, working with one unified model. Figure 2 illustrates the concepts of sufficiency and realizability: the second sufficient class of strategies, that is, Poisson-related strategies, are realizable. Note that working in the space of π-strategies, or in the space of Poisson-related strategies, or even in the general space of all π-ξ-strategies, leads to the same optimal values of the objectives. The details can be found in [22], see also the end of Section 4. The goal of the current paper was to investigate in depth the idea of realizability and emphasize the class of Poisson-related strategies which is simultaneously realizable and sufficient. Moreover, when looking for the best Poisson-related strategy, one can invoke all the methods developed for discrete-time Markov decision processes. Note that the objective functionals are total undiscounted losses, and absolutely no conditions were imposed on the cost and transition rates.

In the framework of games, realizable solutions look problematic. (See [24, Th.10.8, 10.11].)

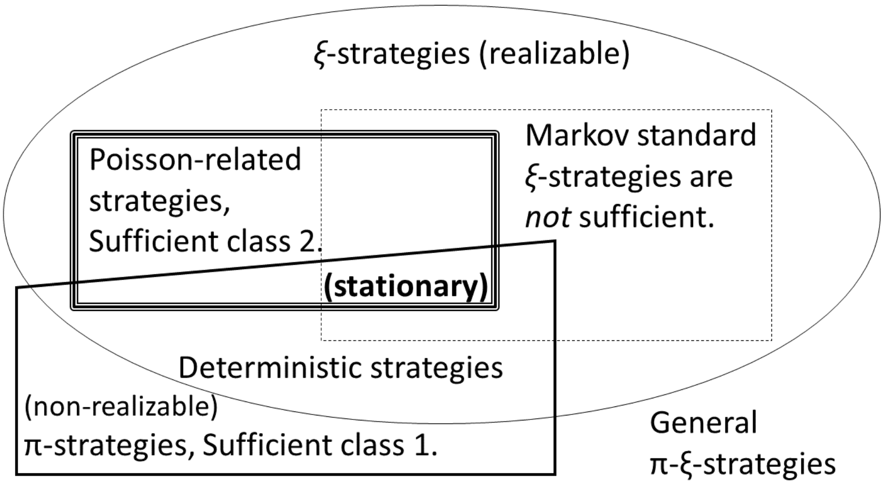
ξ-strategies (realizable)
Poisson-related strategies, Sufficient class 2.
Markov standard ξ-strategies are *not* sufficient.
(stationary)
Deterministic strategies
(non-realizable)
π-strategies, Sufficient class 1.
General π-ξ-strategies

Figure 2: Overview of sufficient and realizable strategies.

7 Appendix

Lemma 1 *Suppose (E,ρ) is a complete separable metric space and let the set $E_d = \{e_1, e_2, \dots\} \subseteq E$ be dense in E. Let P be a probability measure on E. If, for all e_i, $k \in \mathbb{N}$, $P\left(O\left(e_i, \frac{1}{k}\right)\right) \in \{0,1\}$, then P is a Dirac measure. Here $O\left(e_i, \frac{1}{k}\right) = \left\{e \in E : \ \rho(e, e_i) < \frac{1}{k}\right\}$ is an open ball.*

Proof. Collection of the balls $\left\{O\left(e_i, \frac{1}{k}\right) : \ i, k \in \mathbb{N}\right\}$ is a base of the topology in E generated by the metric ρ. The formulated property of the P measure implies that $P(B) \in \{0,1\}$ for every open set $B \subseteq E$. Hence P is a Dirac measure. □

Proof of Theorem 1. We will show that the following statements are equivalent.

(a) A control strategy S is realizable for (h_{n-1}, ξ_n) on the interval $(t_{n-1}, T_n] \neq \emptyset$.

(b) There is a complete probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ and a measurable (with respect to $(u, \tilde{\omega})$) process $A(\cdot)$ on $\mathbb{R}_+ \times \tilde{\Omega}$ with values in $\mathbf{A}$ such that, for almost all $u \in \mathbb{R}_+$, for each $\Gamma_{\mathbf{A}} \in \mathcal{B}(\mathbf{A})$, $\tilde{P}(A(u) \in \Gamma_{\mathbf{A}}) = \pi_n(\Gamma_{\mathbf{A}} | h_{n-1}, \xi_n, u)$ and, for each $\theta \in \mathbb{R}_+$, for every bounded measurable function $\hat{q}(\cdot)$ on $\mathbf{A}$, the integral $\int_{(0,\theta]} \hat{q}(A(u))du$ is degenerate (not random), that is, equals a constant $\tilde{P}$-a.s.

(c) For almost all $u \in \mathbb{R}_+$, $\pi_n(\cdot | h_{n-1}, \xi_n, u) = \delta_{\varphi(u)}(\cdot)$ is a Dirac measure, where $\varphi(\cdot)$ is an $\mathbf{A}$-valued measurable function on $\mathbb{R}_+$.

(d) There is a complete probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ and a measurable (with respect to $(u, \tilde{\omega})$) process $A(\cdot)$ on $\mathbb{R}_+ \times \tilde{\Omega}$ with values in $\mathbf{A}$ such that
– for almost all $u \in \mathbb{R}_+$, for each $\Gamma_{\mathbf{A}} \in \mathcal{B}(\mathbf{A})$, $\tilde{P}(A(u) \in \Gamma_{\mathbf{A}}) = \pi_n(\Gamma_{\mathbf{A}} | h_{n-1}, \xi_n, u)$ and
– for each bounded measurable function $\hat{q}(\cdot)$ on $\mathbf{A}$, the integrals $\int_{I_1} \hat{q}(A(u))du$ and $\int_{I_2} \hat{q}(A(u))du$ are independent for any bounded non-overlapping intervals $I_1, I_2 \subset \mathbb{R}_+$.

Let us show that (a) implies (b).

Suppose the total jump rate $\hat{q}(a)$ is an arbitrary measurable bounded function. Then, according to item (a) of Definition 3, for almost all $u \in \mathbb{R}_+$, $\hat{q}(\pi, u) = \tilde{E}[\hat{q}(A(u, \tilde{\omega}))]$, where $\hat{q}(\pi, u) = \int_{\mathbf{A}} \hat{q}(a) \pi_n(da | h_{n-1}, \xi_n, u)$. Therefore, according to item (b) of Definition 3, the cumulative distribution function of the sojourn time Θ_n, given by

$$\int_{(0,\theta]} G_n^{\xi,\hat{q}}(\mathbf{X}, h_{n-1}, \xi_n, u)du = 1 - e^{-\int_{(0,\theta]} \hat{q}(\pi,u)du} = 1 - e^{-\tilde{E}\left[\int_{(0,\theta]} \hat{q}(A(u,\tilde{\omega}))du\right]} \quad \text{for each } \theta < \infty,$$

must coincide with $\tilde{E}\left[1 - e^{-\int_{(0,\theta]}\hat{q}(A(u,\tilde{\omega}))du}\right]$, that is, we have

$$e^{-\tilde{E}\left[\int_{(0,\theta]}\hat{q}(A(u,\tilde{\omega}))du\right]} = \tilde{E}\left[e^{-\int_{(0,\theta]}\hat{q}(A(u,\tilde{\omega}))du}\right].$$

Since function e^{-z} is strictly convex, we conclude that, for each $\theta \in \mathbb{R}_+$, the integral $\int_{(0,\theta]}\hat{q}(A(u,\tilde{\omega}))$ $\times du$ is not random. Therefore, assertion (a) implies (b).

Let us prove that (b) implies (c).

Suppose $\pi_n(\cdot|h_{n-1},\xi_n,s)$ is not a Dirac measure on a subset of a positive Lebesgue measure, that is, on a subset of a finite interval $(0,\hat{t}] \subset \mathbb{R}_+$ having a positive Lebesgue measure. The goal is to show that assertion (b) is violated. We are going to apply Lemma 1 to $E = \mathbf{A}$, where $\mathbf{A}$ has been equipped with a compatible metric ρ. Below, $O(a,\varepsilon) = \{b \in \mathbf{A} : \rho(a,b) < \varepsilon\}$ is an open ball. If for any $e_m \in E_d$, for any $k \in \mathbb{N}$, the set $\left\{t \in (0,\hat{t}] : \pi_n(O(e_m,\frac{1}{k})|h_{n-1},\xi_n,t) \in (0,1)\right\}$ is null, then the set

$$\left\{t \in (0,\hat{t}] : \exists\, e_m \in E_d,\ \exists\, k \in \mathbb{N} : \pi_n(O(e_m,\frac{1}{k})|h_{n-1},\xi_n,t) \in (0,1)\right\}$$

is also null as a countable union of null sets, and therefore, according to Lemma 1, for almost all $t \in (0,\hat{t}]$, $\pi_n(\cdot|h_{n-1},\xi_n,t)$ is a Dirac measure. From the obtained contradiction, we conclude that there are $\hat{e}_m \in E_d$ and $\hat{k} \in \mathbb{N}$ such that $Leb(\Gamma_{\mathbb{R}}) > 0$, where

$$\Gamma_{\mathbb{R}} = \left\{t \in (0,\hat{t}] : \pi_n(O(\hat{e}_m,\frac{1}{\hat{k}})|h_{n-1},\xi_n,t) \in (0,1)\right\}. \tag{24}$$

Now, suppose assertion (b) is valid.

Consider the function $\hat{q}(a) = I\left\{a \in O(\hat{e}_m,\frac{1}{\hat{k}})\right\}$ and the integrals $V(t) = \int_{(0,t]}\hat{q}(A(u))du$ for $t \in (0,\hat{t}]$, which are non-random if assertion (b) is valid. To be more precise, for each rational $t \in (0,\hat{t}]$, there is a number $f(t)$ such that $\tilde{P}(V(t) = f(t)) = 1$. Hence

$$\tilde{P}(\text{for all rational } t \in (0,\hat{t}]\ V(t) = f(t)) = 1.$$

Since for each $\tilde{\omega} \in \tilde{\Omega}$ the function $V(\cdot)$ is absolutely continuous, we can extend the definition of the function f to the whole interval $(0,\hat{t}]$ in such a way that it is also absolutely continuous: it is sufficient to take an arbitrary $\tilde{\omega}$ such that $V(t) = f(t)$ for all rational $t \in (0,\hat{t}]$ and extend this equality for the whole interval $t \in (0,\hat{t}]$. As a result, $\tilde{P}(\forall t \in (0,\hat{t}]\ V(t) = f(t)) = 1$. Therefore, function $f(\cdot)$ is differentiable everywhere apart from a null set N and

$$\tilde{P}(\hat{\Omega}) = 1, \tag{25}$$

where

$$\hat{\Omega} = \left\{\tilde{\omega} \in \tilde{\Omega} : \hat{q}(A(t)) = h(t) = \frac{df}{dt} \text{ for all } t \in (0,\hat{t}] \setminus N_{\tilde{\omega}}\right\} = \{\tilde{\omega} \in \tilde{\Omega} : \forall t \in (0,\hat{t}]\ V(t) = f(t)\}.$$

Here $N_{\tilde{\omega}} := N \bigcup \{t : \hat{q}(A(t,\tilde{\omega})) \neq h(t)\}$ and $Leb(N_{\tilde{\omega}}) = 0$ for all $\tilde{\omega} \in \tilde{\Omega}$. Below, if necessary, we extend the function $h(\cdot)$ with values in $\{0,1\}$ on the set $\Gamma_{\mathbb{R}} \subset (0,\hat{t}]$, defined in (24), in an arbitrary way. (Remember, $\hat{q}(A(t)) \in \{0,1\}$.) For the set

$$\Gamma = \{(t,\tilde{\omega}) : t \in \Gamma_{\mathbb{R}},\ \hat{q}(A(t,\tilde{\omega})) = h(t)\},$$

we have

$$Leb(\Gamma_{\tilde{\omega}}) = Leb(\{t : t \in \Gamma_{\mathbb{R}},\ \hat{q}(A(t,\tilde{\omega})) = h(t)\}) = Leb(\Gamma_{\mathbb{R}} \setminus N_{\tilde{\omega}}) = Leb(\Gamma_{\mathbb{R}})$$

for all $\tilde{\omega} \in \hat{\Omega}$. Therefore, (25) implies that

$$Leb \times \tilde{P}(\Gamma) = Leb(\Gamma_{\mathbb{R}}) > 0. \tag{26}$$

On the other hand, according to (24), since for almost all $t \in \Gamma_{\mathbb{R}}$,

$$\tilde{P}(\hat{q}(A(t)) = 1) = \tilde{P}\left(A(t) \in O\left(\hat{e}_m, \frac{1}{\hat{k}}\right)\right) = \pi_n\left(O\left(\hat{e}_m, \frac{1}{\hat{k}}\right)|h_{n-1}, \xi_n, t\right) \in (0,1),$$

and similarly $\tilde{P}(\hat{q}(A(t)) = 0) \in (0,1)$, we have inequality $0 < Leb \times \tilde{P}(\Gamma) < Leb(\Gamma_{\mathbb{R}})$ because $h(\cdot) \in \{0,1\}$. The obtained contradiction confirms that assertion (b) is violated.

We have proved that (b) implies (c).

If statement (c) holds, then one can take $\tilde{\Omega} = \{\tilde{\omega}\}$ as a singleton with the trivial σ-algebra and the trivial probability $\tilde{P}(\tilde{\Omega}) = 1$. After we put $A(s, \tilde{\omega}) = \varphi(s)$, we see that statement (a) holds, as well as statement (d). (Remember, any two random variables on the trivial probability space are independent.)

Now suppose statement (d) is valid. For an arbitrary fixed bounded measurable function $\hat{q}$ on $\mathbf{A}$ and fixed $\theta \in \mathbb{R}_+$,

$$\int_{(0,\theta]} \hat{q}(A(u))du = \sum_{i=1}^{k} \int_{(t_i,t_{i+1}]} \hat{q}(A(u))du,$$

where $t_i = \frac{(i-1)\theta}{k}$ and $k \in \mathbb{N}$ is a fixed number. Since the integrals $\int_{(t_i,t_{i+1}]} \hat{q}(A(u))du$ are independent from each other, the variance of $\int_{(0,\theta]} \hat{q}(A(u))du$ with respect to $\tilde{P}$ satisfies equality

$$Var\left(\int_{(0,\theta]} \hat{q}(A(u))du\right) = \sum_{i=1}^{k} Var\left(\int_{(t_i,t_{i+1}]} \hat{q}(A(u))du\right).$$

However,

$$Var\left(\int_{(t_i,t_{i+1}]} \hat{q}(A(u))du\right) \le \tilde{E}\left[\left(\int_{(t_i,t_{i+1}]} \hat{q}(A(u))du\right)^2\right] \le \left(\sup_{a\in\mathbf{A}} |\hat{q}(a)|\frac{\theta}{k}\right)^2,$$

so that

$$Var\left(\int_{(0,\theta]} \hat{q}(A(u))du\right) \le \frac{\theta^2 \left(\sup_{a\in\mathbf{A}} |\hat{q}(a)|\right)^2}{k}.$$

Since this inequality is valid for each $k \in \mathbb{N}$,

$$Var\left(\int_{(0,\theta]} \hat{q}(A(u))du\right) = 0$$

and the integral $\int_{(0,\theta]} \hat{q}(A(u))du$ is not random. Statement (b) holds true. The proof is competed. □

References

[1] Altman, E. *Constrained Markov Decision Processes.* Chapman and Hall/CRC, Boca Raton, 1999.

[2] Bellman, R. *Dynamic Programming.* Princeton University Press, Princeton, 1957.

[3] Bertsekas, D. and Shreve, S. *Stochastic Optimal Control.* Academic Press, NY, 1978.

[4] Dufour, F. and Piunovskiy, A. The expected total cost criterion for Markov decision processes under constraints. *Adv. Appl. Prob.* **45** (2013) 837-859.

[5] Dufour, F. and Prieto-Rumeau, T. Conditions for the solvability of the linear programming formulation for constrained discounted Markov decision processes. *Appl Math. Optim.* (2016) DOI 10.1007/s00245-015-9307-3.

[6] Feinberg, E.A. and Shwartz, A. Constrained discounted dynamic programming. *Math. Oper. Res.* **21** (1996) 922-945.

[7] Feinberg, E. Total reward criteria. In *Handbook of Markov Decision Processes.* (E.Feinberg and A.Shwartz ed.), Kluwer, Boston/Dordrecht/London, 2002, 173-207.

[8] Feinberg, E. Continuous time discounted jump Markov decision processes: a discrete-event approach. *Math. Oper. Res.* **29** (2004) 492-524.

[9] Feinberg, E. Reduction of discounted continuous-time MDPs with unbounded jump and reward rates to discrete-time total-reward MDPs. In *Optimization, Control, and Applications of Stochastic Systems* (D.Hernandez-Hernandez and J.A.Minjares-Sosa ed.), Birkhauser, 2012, 77-97.

[10] Guo, X., Hernández-Lerma, O. and Prieto-Rumeau, T. A survey of recent results on continuous-time Markov decision processes. *Top*, **14** (2006) 177-257.

[11] Guo, X. and Hernández-Lerma, O. *Continuous-Time Markov Decision Processes: Theory and Applications.* Springer-Verlag, Heidelberg, 2009.

[12] Guo, X. and Piunovskiy, A. Discounted continuous-time Markov decision processes with constraints: unbounded transition and loss rates. *Math. Oper. Res.* **36** (2011) 105-132.

[13] Guo, X., Huang, Y. and Zhang, Y. Constrained continuous-time Markov decision process3es on the finite horizon. *Appl Math. Optim.* (2016) DOI 10.1007/s00245-016-9352-6.

[14] Hernández-Lerma, O. and Lasserre, J.B. *Further Topics on Discrete-Time Markov Control Processes.* Springer-Verlag, NY, 1999.

[15] Howard, R.A. *Dynamic Programming and Markov Processes.* M.I.T Press, Cambridge, 1960.

[16] Jacod, J. Multivariate point processes: predictable projection, Radon-Nykodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie verw. Gebite.* **31** (1975) 235-253.

[17] Kallianpur, G. *Stochastic Filtering Theory.* Springer, New York, 1980.

[18] Kitaev, M. Semi-Markov and jump Markov controlled models: average cost criterion. *Theory. Probab. Appl.* **30** (1986) 272-288.

[19] Piunovskiy, A. *Optimal Control of Random Sequences in Problems with Constraints.* Kluwer, Dordrecht, 1997.

[20] Piunovskiy, A. and Zhang, Y. Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach. *SIAM J. Control Optim.* **49** (2011) 2032-2061.

[21] Piunovskiy, A. *Examples in Markov Decision Processes.* Imperial College Press, London, 2013.

[22] Piunovskiy, A. Randomized and relaxed strategies in continuous-time Markov decision processes. *SIAM J. Control Optim.*, **53** (2015) 3503-3533.

[23] Piunovskiy, A. Sufficient classes of strategies in continuous-time Markov decision processes with total expected cost. In *Modern Trends in Controlled Stochastic Processes, V.II* (A.Piunovskiy ed.) Luniver Press, Frome, 2015, 190-212.

[24] Prieto-Rumeau, T. and Hernandez-Lerma, O. *Selected Topics on Continuous-Time Controlled Markov Chains and Markov Games.* Imperial College Press, London, 2012.

[25] Tijms, H.C. *A First Course in Stochastic Models.* Wiley, Chichester, 2003.

[26] Yushkevich, A. Controlled Markov models with countable state space and continuous time. *Theory Probab. Appl.* **22** (1977) 215-235.

[27] Yushkevich, A. On reducing a jump controllable Markov model to a model with discrete time. *Theory Probab. Appl.* **25** (1980) 58-69.

[28] Yushkevich, A. Controlled jump Markov models. *Theory Probab. Appl.* **25** (1980) 244-266.