Optimal Control of Probability on a Target Set for Continuous-Time Markov Chains

Chenglin Ma and Huaizhong Zhao

***Abstract*— In this article, a stochastic optimal control problem is considered for a continuous-time Markov chain taking values in a denumerable state space over a fixed finite horizon. The optimality criterion is the probability that the process remains in a target set before and at a certain time. The optimal value is a superadditive capacity of target sets. Under some minor assumptions for the controlled Markov process, we establish the dynamic programming principle, based on which we prove that the value function is a classical solution of the Hamilton-Jacobi-Bellman (HJB) equation on a discrete lattice space. We then prove that there exists an optimal deterministic Markov control under the compactness assumption of control domain. We further prove that the value function is the unique solution of the HJB equation. We also consider the case starting from the outside of the target set and give the corresponding results. Finally, we apply our results to two examples.**

***Index Terms*— Controlled Markov chains, dynamic programming principle (DPP), Hamilton-Jacobi-Bellman (HJB) equation, optimal controls, risk probability criteria.**

I. Introduction

Stochastic optimal control problems for Markov chains, also known as Markov decision processes (MDPs), have been widely studied due to their rich applications in real-world contexts, such as in communication engineering [1], finance [6], queuing systems [21], control of epidemics [24] and so on. Existing articles mainly focus on MDPs with expected/average reward criteria. See, for example, [7], [9], [12], [17], [22], [25], and [26]. However, such a setup is not always suitable in some applications. For example, when we measure the market risk in the areas of finance and economics, it is reasonable to minimize the probability of loss exceeding a fixed value. Inspired by the considerations of real-world contexts, some authors started to study MDPs with risk probability criteria.

MDPs with risk probability criterions can be roughly divided into two kinds: the discrete-time case and the continuous-time case. For the discrete-time scenario, a general study can be found in [5] and [29]. Recently, a discrete-time optimal dividend problem with risk probability criteria has been considered in [28], the aim of which was to minimize the risk probability of reaching a given dividend goal before the time of ruin and find the optimal dividend policy. In [13], a two-player nonzero-sum discrete-time stochastic games under probability criterion was considered, and it was shown that the optimal value function for each player is the unique solution to the corresponding optimality equation, and the existence of Nash equilibria was established under mild conditions.

The continuous-time MDPs with risk probability criteria were considered for the first time in [16]. Under some conditions, it was proved that the value function is a solution to the optimality equation. Following the publication of this work, there were some works on continuous-time MDPs with risk probability criteria, such as [4] and [15]. Bhabak and Saha [4] studied a zero-sum stochastic game for continuous-time Markov chains. Under some assumptions, they showed the existence of value of the game and also characterized it as the unique solution of a pair of Shapley equations. Huo and Guo [15] dealt with finite horizon continuous-time MDPs with unbounded transition rates and established the existence and uniqueness of a solution of the corresponding optimality equation. They also proved the existence of a risk probability optimal policy.

In this article, we would like to find the optimal control processes to maximize the probability that the controlled Markov process is always in a target set during the fixed finite horizon $[0, T]$. Such a risk probability setup can be regarded as the surviving probability on a safety set in many real-world contexts, such as the number of cancer cells in a patient in a certain safety range. In the context of this article, the admissible controls, we consider, are processes taking values in a compact control domain and being adapted to the natural filtration generated by the underlying Markov chain. We firstly give the dynamic programming principle (DPP) by considering a family of stochastic optimal control subproblems initiated by different times and states. We find that the global optimal control is also locally optimal over any second half-horizon $[t, T]$ in the sense of conditional expectation. We then establish the relationship among these subproblems by deriving the so-called Hamilton-Jacobi-Bellman (HJB) equation. This is a nonlinear first-order differential-difference equation. The value function is a classical solution of the HJB equation due to its right differentiability with respect to (w.r.t.) the time variable. By the compactness assumption of control domain, we give the existence theorem of optimal deterministic Markov controls for the dynamic programming (DP) problem by employing measurable selection theorem (see [2] and references therein). We further prove that the value function is the unique solution of the HJB equation. We then also consider the case starting from the outside of the target set to maximize the probability on the target set from any time $t_0 \in (0, T]$.

A. Problem statement

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space on which a continuous-time Markov chain $\{X_t, 0 \leq t \leq T\}$ is defined over finite horizon $[0, T]$ for a fixed $T > 0$. We denote by $\{\mathcal{F}_t, 0 \leq t \leq T\}$ the natural filtration generated by $X(\cdot)$ and augmented by all P-null sets of $\mathcal{F}$, that is, $\mathcal{F}_t = \sigma\{X_s, 0 \leq s \leq t\} \vee \mathcal{N}_p, \ 0 \leq t \leq T$, where $\mathcal{N}_p$ is the set of all P-null set of $\mathcal{F}$. The state space S of the process X_t is a denumerable space endowed with a discrete topology. A finite-state subset $B \subset S$ is the target set with $B^c := S \setminus B$. The control domain $U \subset \mathbf{R}$ is a nonempty compact set equipped with

Manuscript received 14 December 2022; accepted 13 May 2023. This work was supported in part by the National Key R&D Programme of China under Grant 2020YFA0712400 and in part by the Engineering and Physical Sciences Researvh Council under Grant EP/S005293/2. Recommended by Associate Editor S.-J. Chung. (Corresponding author: Huaizhong Zhao)

Chenglin Ma is with the School of Mathematics, Shandong University, Jinan 250100, China (e-mail: machenglin@mail.sdu.edu.cn).

Huaizhong Zhao is with the Department of Mathematical Sciences, Durham University, DH1 3LE Durham, U.K., and also with the Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, Qingdao 266237, China (e-mail: huaizhong.zhao@durham.ac.uk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2023.3278789.

Digital Object Identifier 10.1109/TAC.2023.3278789

Borel σ-algebra $\mathcal{B}(U)$. Let $\mathcal{U}$ denote the *admissible* control set

$$\mathcal{U} := \{u : [0,T] \times \Omega \to U | u \text{ is } \{\mathcal{F}_t\}\textit{-adapted}\}.$$

An admissible control $u(\cdot)$ is called a deterministic Markov control, if the value of $u(\cdot)$ only depends on the current time and state. Denote by $\mathcal{M}$ the set of all deterministic Markov controls over $[0,T]$. Define $\mathcal{F}_s^t = \sigma\{X_r^{t,x}, r \in [t,s]\} \vee \mathcal{N}_p$; the admissible control set $\mathcal{U}_t$ consists of processes taking values in U and being adapted to $\{\mathcal{F}_s^t\}$, and the admissible control set $\mathcal{M}_t$ consists of deterministic Markov controls over $[t,T]$. Obviously, $\mathcal{M} \subset \mathcal{U}$ and $\mathcal{M}_t \subset \mathcal{U}_t$, for all $0 < t \leq T$.

For any given $u(\cdot) \in \mathcal{U}$, the process $X_{t,u(\cdot)}$ is assumed to satisfy the regularity condition: $\lim_{s \downarrow t} P\{X_{s,u(\cdot)}^{t,x} = x'\} = \delta_{xx'}$, where $\delta_{xx'} = 1$ if $x' = x$ or 0 otherwise, which implies the process $X_{t,u(\cdot)}$ has only finitely many jumps with probability one over the finite horizon $[0,T]$. The superscript of $X_{s,u}^{t,x}$ denotes the initial time and state, we consider, and $X_{s,u}^{0,x_0}$ can be simplified as $X_{s,u}^{x_0}$.

For each $u(\cdot) \in \mathcal{U}_t$ (with $u_t = u \in U$), the infinitesimal transition probabilities of $X_{t,u(\cdot)}$ are given by

$$P\{X_{t+\delta,u(\cdot)}^{t,x} = x'\} = \begin{cases} \lambda_{xx'}(t,u)\delta + o(\delta), \ \textit{if } x' \neq x, \\ 1 + \lambda_{xx}(t,u)\delta + o(\delta), \ \textit{otherwise}, \end{cases} \tag{1}$$

where $\lambda_{ij}(t,u)$ are transition rates of X_t and supposed to satisfy the following assumptions.

1) If $j \neq i$, $\lambda_{ij}(t,u) \geq 0$, for any $(t,u) \in [0,T] \times U$.

2) The transition rates are conservative, that is, for any $(t,u) \in [0,T] \times U$ and $i \in S$, we have $\sum_{j \in S} \lambda_{ij}(t,u) = 0$.

3) The transition rates are stable, that is, for any $(t,u) \in [0,T] \times U$, we have $\sup_{i \in S} |\lambda_{ii}(t,u)| < \infty$.

4) The transition rates $\lambda_{ij}(\cdot,\cdot)$ are continuous in $[0,T] \times U$ for any $i,j \in S$.

Remark 1: 1). All the properties of the controlled Markov process are determined by the transition rates, so it is sufficient to make assumptions about the transition rates only. However, in practice, it is a difficult task to identify the transition rates; statistical analysis is useful in this context. This is not the aim of this article, so we will not expand this aspect here and leave it for a future project.

2). For each fixed $i \in S$, the transition rate λ_{ij} is bounded due to the continuity and stability.

Problem (S): The optimal control problem we are interested in is to maximize the *utility* functional given by the probability staying on the given target set B: for $(t,x) \in [0,T] \times B$,

$$J(t,x,u(\cdot)) := P\{X_{s,u(\cdot)}^{t,x} \in B, \ \forall s \in [t,T]\} \tag{2}$$

over $u(\cdot) \in \mathcal{U}_t$. The value function associated with (2) is defined as

$$V(t,x) := \sup_{u(\cdot) \in \mathcal{U}_t} J(t,x,u(\cdot)), \quad (t,x) \in [0,T] \times B. \tag{3}$$

Note the boundary condition

$$\begin{cases} V(T,x) = 1, \ \forall x \in B, \\ V(t,x) = 0, \ \forall (t,x) \in [0,T] \times B^c. \end{cases} \tag{4}$$

Remark 2: 1) Let $\tau_1 := \inf\{s \geq t, X_s \neq x | X_t = x\}$ denote the first jump time of X_s after t. By (1), for any control $u(\cdot) \in \mathcal{M}_t$, we have

$$P\{\tau_1 > T\} = \exp\{\int_t^T \lambda_{xx}(r, u(r))dr\} > 0 \tag{5}$$

which implies that $V(t,x) > 0$ always holds for each $x \in B$.

2). For a fixed $x \in B$, let $\tau(u(\cdot)) := \inf\{s \geq t, X_{s,u(\cdot)}^{t,x} \notin B\}$ denote the first exit time of $X_{s,u(\cdot)}^{t,x}$ from B; then, the utility functional (2) can be rewritten as

$$J(t,x,u(\cdot)) = P\{\tau(u(\cdot)) > T\}.$$

Definition 1: A control $u^*(\cdot) \in \mathcal{U}_t$ is called an optimal control of Problem (S) if $u^*(\cdot)$ is the control such that

$$V(t,x) = J(t,x,u^*(\cdot)). \tag{6}$$

Remark 3: The optimality criterion in this article is similar to that in [14] on optimal risk probability for first passage models of semi-MDPs, since there is no reward/cost structure in these models. The admissible controls we consider here are stochastic processes adapted to the natural filtration generated by the underlying Markov chain, which are different from the policies studied in [14] as well as some continuous-time MDPs with risk probability criteria, such as [15], [16]. In these works, the policies are only taken at each jump point.

II. Dynamic programming principle

One of the most commonly used approaches to solve stochastic optimal control problems is to establish the DPP based on the pioneering work of Bellman [3]. The basic idea of DPP is to consider a family of optimal control subproblems initiated at different times and states. Then, the next step is to establish the connections among these subproblems and solve all of them finally. However, for any $s \in (t,T]$, $X_s^{t,x}$ is a random variable in $(\Omega, \mathcal{F}, P)$ rather than a deterministic state in S, but it can be regarded as almost surely deterministic under the conditional probability measure $P\{\cdot|\mathcal{F}_s^t\}(\omega)$ for each fixed $\omega \in \Omega$, in the sense that all the dynamics of $X^{t,x}$ during the time period $[t,s]$ are known under the filtration $\mathcal{F}_s^t$, as explained in [31]. Then for any $s \in (t,T]$ and a given $u(\cdot) \in \mathcal{U}_t$, we have

$$J(s, X_s(\omega), u(\cdot)) = P\{X_{r,u(\cdot)}^{s,X_{s,u(\cdot)}^{t,x}} \in B, \forall r \in [s,T] | \mathcal{F}_s^t\}(\omega), P - a.s.$$

Then, for each $(t,x) \in [0,T) \times B$ and $s \in (t,T]$, by taking conditional expectation, we have

$$\begin{aligned} J(t,x,u) =& E\left[I_{\{X_{r,u(\cdot)}^{t,x} \in B, \forall r \in [t,T]\}}\right] \\ =& E\left[I_{\{X_{r,u(\cdot)}^{t,x} \in B, \forall r \in [t,s]\}} E[I_{\{X_{r,u(\cdot)}^{t,x} \in B, \forall r \in [s,T]\}} | \mathcal{F}_s^t]\right] \\ =& E\left[I_{\{X_{r,u(\cdot)}^{t,x} \in B, \forall r \in [t,s]\}} E[I_{\{X_{r,u(\cdot)}^{s,X_s} \in B, \forall r \in [s,T]\}} | \mathcal{F}_s^t]\right] \\ =& E\left[I_{\{X_{r,u(\cdot)}^{t,x} \in B, \forall r \in [t,s]\}} J(s, X_{s,u(\cdot)}^{t,x}, u(\cdot))\right], \end{aligned} \tag{7}$$

where I_F denotes the indicator function of set $F \in \mathcal{F}$. Here we used the flow relation $X_{r,u(\cdot)}^{t,x} = X_{r,u(\cdot)}^{s,X_{s,u(\cdot)}^{t,x}}$, $P\{\cdot \mid \mathcal{F}_s^t\}(\omega) - a.s.$, for all $t \leq s \leq r \leq T$ and each fixed $\omega \in \Omega$.

Theorem 1: For any $(t,x) \in [0,T) \times B$ and $s \in (t,T]$, the value function $V(t,x)$ satisfies the DP equation

$$V(t,x) = \sup_{u(\cdot) \in \mathcal{U}_t} E\left[I_{\{X_{r,u(\cdot)}^{t,x} \in B, \forall r \in [t,s]\}} V(s, X_{s,u(\cdot)}^{t,x})\right]. \tag{8}$$

In particular, for a sufficiently small $\delta > 0$, we have

$$V(t,x) = \sup_{u(\cdot) \in \mathcal{U}_t} E\left[V(t+\delta, X_{t+\delta,u(\cdot)}^{t,x})\right] + o(\delta). \tag{9}$$

Proof: First, for any $\varepsilon > 0$, it is easy to know that there exists $\hat{u}(\cdot) \in \mathcal{U}_t$ such that for any $t \leq s \leq T$,

$$\begin{aligned} &\sup_{u(\cdot) \in \mathcal{U}_t} E\left[I_{\{X_{r,u(\cdot)}^{t,x} \in B, \forall r \in [t,s]\}} V(s, X_{s,u(\cdot)}^{t,x})\right] - \varepsilon \\ <& E\left[I_{\{X_{r,\hat{u}(\cdot)}^{t,x} \in B, \forall r \in [t,s]\}} V(s, X_{s,\hat{u}(\cdot)}^{t,x})\right] \\ =& \sum_{x' \in S} V(s,x') E\left[I_{\{X_{r,\hat{u}(\cdot)}^{t,x} \in B, \forall r \in [t,s]\}} I_{\{X_{s,\hat{u}(\cdot)}^{t,x} = x'\}}\right]. \end{aligned} \tag{10}$$

Note, for each $x' \in S$, there exists a control $u^{sx'}(\cdot) \in \mathcal{U}_s$ such that

$$V(s,x') - \varepsilon < J(s,x',u^{sx'}(\cdot)). \tag{11}$$

Let

$$\tilde{u}(\cdot) := \begin{cases} \hat{u}_r(\cdot), & r \in [t,s), \\ \sum_{x'} u_r^{sx'}(\cdot) I_{\{X_s = x'\}}, & r \in [s,T], \end{cases}$$

which forms an admissible control in $\mathcal{U}_t$. Thus, by (7) and combing the 2ε's in (10) and (11), we have

$$\begin{aligned} &\sup_{u(\cdot)\in\mathcal{U}_t} E[I_{\{X^{t,x}_{r,u(\cdot)}\in B,\forall r\in[t,s]\}} V(s,X^{t,x}_{s,u(\cdot)})] - 2\varepsilon \\ &< \sum_{x'\in S} J(s,x',u^{sx'}(\cdot)) E[I_{\{X^{t,x}_{r,\hat{u}(\cdot)}\in B,\forall r\in[t,s]\}} I_{\{X^{t,x}_{s,\hat{u}(\cdot)}=x'\}}] \\ &= E[I_{\{X^{t,x}_{r,\tilde{u}(\cdot)}\in B,\forall r\in[t,s]\}} J(s,X^{t,x}_{s,\tilde{u}(\cdot)},\tilde{u}(\cdot))] \\ &= J(t,x,\tilde{u}(\cdot)) \le V(t,x). \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we have

$$V(t,x) \ge \sup_{u(\cdot)\in\mathcal{U}_t} E[I_{\{X^{t,x}_{r,u(\cdot)}\in B,\forall r\in[t,s]\}} V(s,X^{t,x}_{s,u(\cdot)})]. \tag{12}$$

Conversely, for any $\varepsilon > 0$, there exists $u^{\varepsilon}(\cdot) \in \mathcal{U}_t$ such that

$$\begin{aligned} &-\varepsilon + V(t,x) < J(t,x,u^{\varepsilon}(\cdot)) \\ &= E[I_{\{X^{t,x}_{r,u^{\varepsilon}(\cdot)}\in B,\forall r\in[t,s]\}} J(s,X^{t,x}_{s,u^{\varepsilon}(\cdot)},u^{\varepsilon}(\cdot))] \\ &\le E[I_{\{X^{t,x}_{r,u^{\varepsilon}(\cdot)}\in B,\forall r\in[t,s]\}} V(s,X^{t,x}_{s,u^{\varepsilon}(\cdot)})] \\ &\le \sup_{u(\cdot)\in\mathcal{U}_t} E[I_{\{X^{t,x}_{r,u(\cdot)}\in B,\forall r\in[t,s]\}} V(s,X^{t,x}_{s,u(\cdot)})]. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we deduce that

$$V(t,x) \le \sup_{u(\cdot)\in\mathcal{U}_t} E[I_{\{X^{t,x}_{r,u(\cdot)}\in B,\forall r\in[t,s]\}} V(s,X^{t,x}_{s,u(\cdot)})]. \tag{13}$$

Combining (12) and (13), we thus obtain (8).

To see (9), from (8), for a sufficiently small $\delta > 0$, we have

$$V(t,x) = \sup_{u(\cdot)\in\mathcal{U}_t} E[I_{\{X^{t,x}_{r,u(\cdot)}\in B,\forall r\in[t,t+\delta]\}} V(t+\delta,X^{t,x}_{t+\delta,u(\cdot)})].$$

However, the transition probabilities (1) imply that the probability of two jumps occurring in a small interval of length δ is $o(\delta)$. In fact,

$$\begin{aligned} &E[I_{\{X^{t,x}_{s,u(\cdot)}\in B,\forall s\in[t,t+\delta]\}} V(t+\delta,X^{t,x}_{t+\delta,u(\cdot)})] \\ &= \sum_{x'\in B} V(t+\delta,x') E[I_{\{X^{t,x}_{s,u(\cdot)}\in B,\forall s\in[t,t+\delta]\}} I_{\{X^{t,x}_{t+\delta,u(\cdot)}=x'\}}]. \end{aligned}$$

However, the event $\{X^{t,x}_{t+\delta,u(\cdot)} = x' \in B\}$ includes two distinct cases: Ω_1:=$\{X_s$ always stays in B during the period $[t,t+\delta]$ and $X^{t,x}_{t+\delta,u(\cdot)} = x'\}$ and Ω_2:=$\{X_s$ jumps out of B and then returns back with $X^{t,x}_{t+\delta,u(\cdot)} = x' \in B\}$. From (1), we have $P(\Omega_2) = O(\delta^2)$. Then,

$$\begin{aligned} &E[I_{\{X^{t,x}_{t+\delta,u(\cdot)}=x'\}}] \\ &= P(X^{t,x}_{t+\delta,u(\cdot)} = x') = P(\Omega_1) + P(\Omega_2) \\ &= E[I_{\{X^{t,x}_{s,u(\cdot)}\in B,\forall s\in[t,t+\delta]\}} I_{\{X^{t,x}_{t+\delta,u(\cdot)}=x'\}}] + P(\Omega_2) \\ &= E[I_{\{X^{t,x}_{s,u(\cdot)}\in B,\forall s\in[t,t+\delta]\}} I_{\{X^{t,x}_{t+\delta,u(\cdot)}=x'\}}] + o(\delta). \end{aligned}$$

That is,

$$\begin{aligned} &E[I_{\{X^{t,x}_{s,u(\cdot)}\in B,\forall s\in[t,t+\delta]\}} V(t+\delta,X^{t,x}_{t+\delta,u(\cdot)})] \\ &= \sum_{x'\in B} V(t+\delta,x') E[I_{\{X^{t,x}_{t+\delta,u(\cdot)}=x'\}}] + o(\delta). \end{aligned}$$

Then, by the boundary condition (4) of $V(t,x)$, we have

$$\begin{aligned} V(t,x) &= \sup_{u(\cdot)\in\mathcal{U}_t} E[I_{\{X^{t,x}_{r,u(\cdot)}\in B,\forall r\in[t,t+\delta]\}} V(t+\delta,X^{t,x}_{t+\delta,u(\cdot)})] \\ &= \sup_{u(\cdot)\in\mathcal{U}_t} \sum_{x'\in B} V(t+\delta,x') E[I_{\{X^{t,x}_{t+\delta,u(\cdot)}=x'\}}] + o(\delta) \\ &= \sup_{u(\cdot)\in\mathcal{U}_t} \sum_{x'\in S} V(t+\delta,x') E[I_{\{X^{t,x}_{t+\delta,u(\cdot)}=x'\}}] + o(\delta) \\ &= \sup_{u(\cdot)\in\mathcal{U}_t} E[V(t+\delta,X^{t,x}_{t+\delta,u(\cdot)})] + o(\delta). \end{aligned}$$

This completes the proof. ∎

Remark 4: The DP equations we obtained in Theorem 1 are different from the optimality equations given in [15] and [16]. The latter describes the relationship of value functions at successive jump points, but the DP equations here describe the relationship of value functions at any different time points. Furthermore, a nonlinear partial differential equation, i.e., the HJB equation, can be obtained in the next section from the DPP. This was not obtained from optimality equations in [15] and [16]. Our results of the DP equation and the HJB equation for this kind of problem are new.

Remark 5: If $T = \infty$, we can also obtain a DP function, in which the freely chosen time s should be replaced by some stopping times. Some examples of DPP involving stopping times can be found in [18], [30], and references therein.

Theorem 2: If $u^*(\cdot)$ is an optimal control of Problem (S), then for any $s \in (t,T]$

$$V(s,X^{t,x}_{s,u^*(\cdot)}) = J(s,X^{t,x}_{s,u^*(\cdot)},u^*(\cdot)),\ P-a.s., \tag{14}$$

conditional on $\{X^{t,x}_{r,u^*(\cdot)} \in B, \forall r \in [t,s]\}$.

Furthermore, DP equation (8) turns out to be optimal equation

$$V(t,x) = E[I_{\{X^{t,x}_{r,u^*(\cdot)}\in B,\forall r\in[t,s]\}} V(s,X^{t,x}_{s,u^*(\cdot)})], \tag{15}$$

and (9) turns out to be

$$V(t,x) = E\left[V(t+\delta,X^{t,x}_{t+\delta,u^*(\cdot)})\right] + o(\delta). \tag{16}$$

Proof: By (6), (7) and (8), for any $s \in (t,T]$, we have

$$\begin{aligned} V(t,x) &= E[I_{\{X^{t,x}_{r,u^*(\cdot)}\in B,\forall r\in[t,s]\}} J(s,X^{t,x}_{s,u^*(\cdot)},u^*(\cdot))] \\ &\le E[I_{\{X^{t,x}_{r,u^*(\cdot)}\in B,\forall r\in[t,s]\}} V(s,X^{t,x}_{s,u^*(\cdot)})] \\ &\le V(t,x). \end{aligned}$$

Hence indeed,

$$\begin{aligned} &E[I_{\{X^{t,x}_{r,u^*(\cdot)}\in B,\forall r\in[t,s]\}} J(s,X^{t,x}_{s,u^*(\cdot)},u^*(\cdot))] \\ &= E[I_{\{X^{t,x}_{r,u^*(\cdot)}\in B,\forall r\in[t,s]\}} V(s,X^{t,x}_{s,u^*(\cdot)})]. \end{aligned}$$

From Remark 2, $P\{X^{t,x}_{r,u^*(\cdot)} \in B, \forall r \in [t,s]\} > 0$, and by the definitions of utility functional and value function, one has $J(s,X^{t,x}_{s,u^*(\cdot)},u^*(\cdot)) \le V(s,X^{t,x}_{s,u^*(\cdot)})$. It follows that

$$J(s,X^{t,x}_{s,u^*(\cdot)},u^*(\cdot)) = V(s,X^{t,x}_{s,u^*(\cdot)}),\quad P-a.s.,$$

conditional on $\{X^{t,x}_{r,u^*(\cdot)} \in B, \forall r \in [t,s]\}$. Naturally, the DP equations (8) and (9) turn out to be (15) and (16), respectively. ∎

If $u^*(\cdot) \in \mathcal{U}_t$ is an optimal control of problem (S), Theorem 2 means that $u^*(\cdot)$ restricted on $[s,T]$ is also optimal *P-almost surely* for the optimal control subproblem over $[s,T]$ initiated by $(s,X^{t,x}_{s,u^*(\cdot)})$ conditional on $\{X^{t,x}_{r,u^*(\cdot)} \in B, \forall r \in [t,s]\}$ for any $t \le s \le T$.

III. HJB Equation and Existence of Optimal Control

From the DPP obtained in the last section, we will establish the relationships among these optimal control subproblems by deriving the so-called HJB equation in this section. In the classical DP problems, the state processes are usually characterized as stochastic differential equations driven by standard Brownian motions, which implies that the state processes are continuous and the HJB equations are normally backward partial differential equations with continuous terminal conditions, readers are referred to [12], [31] and references therein for more details. Different from the classical HJB equation, the HJB equation, derived below (20) in the context of this article, is a backward differential-difference equation with Dirichlet boundary value and terminal value given by an indicator function of the surviving set. We first discuss the infinitesimal generator of the controlled Markov jump process.

Define the operator L as the infinitesimal generator of X_t. For any $(t,x) \in [0,T) \times S$, $u(\cdot) \in \mathcal{U}_t$ (with $u_t = u \in U$) and bounded function f, from (1) we have

$$\begin{aligned} L^u f(x) &= \lim_{\delta\to 0^+} \frac{1}{\delta} E[f(X^{t,x}_{t+\delta,u(\cdot)}) - f(x)] \\ &= \lim_{\delta\to 0^+} \frac{1}{\delta} \sum_{x'\neq x} P\{X^{t,x}_{t+\delta,u(\cdot)} = x'\}(f(x') - f(x)) \\ &= \sum_{x'\neq x} \lambda_{xx'}(t,u)(f(x') - f(x)). \end{aligned} \tag{17}$$

For each deterministic Markov control $u(\cdot) \in \mathcal{M}_t$, denote $M^f_s := f(X^{t,x}_{s,u(\cdot)}) - \int_t^s L^{u(\cdot)} f(X^{t,x}_{r,u(\cdot)})dr$. Then M^f_s is $\{\mathcal{F}^t_s\}$-adapted and integrable since the transition rates of the process X_t is assumed to be stable. Moreover, for $\hat{s} \in [t,s]$, from the definition of infinitesimal generator, it is easy to see

$$\begin{aligned} L^u f(X^{t,x}_{r,u(\cdot)}) &= \lim_{\delta\to 0^+} \frac{1}{\delta} E[f(X^{t,x}_{r+\delta,u(\cdot)}) - f(X^{t,x}_{r,u(\cdot)}) \mid \mathcal{F}^t_r] \\ &= E[\frac{d}{dr} f(X^{t,x}_{r,u(\cdot)}) \mid \mathcal{F}^t_r],\ P-a.s. \end{aligned}$$

It turns out that

$$\begin{aligned} &E[f(X^{t,x}_{s,u(\cdot)}) - f(X^{t,x}_{\hat{s},u(\cdot)}) - \int_{\hat{s}}^s L^{u(\cdot)} f(X^{t,x}_{r,u(\cdot)})dr | \mathcal{F}^t_{\hat{s}}] \\ &= 0,\ P-a.s. \end{aligned}$$

Therefore, for any $t \le \hat{s} \le s \le T$, we have

$$\begin{aligned} E[M^f_s | \mathcal{F}^t_{\hat{s}}] &= E[f(X^{t,x}_{s,u(\cdot)}) - \int_t^s L^{u(\cdot)} f(X^{t,x}_{r,u(\cdot)})dr | \mathcal{F}^t_{\hat{s}}] \\ = -\int_t^{\hat{s}} &L^{u(\cdot)} f(X^{t,x}_{r,u(\cdot)})dr \\ &+ E[f(X^{t,x}_{s,u(\cdot)}) - \int_{\hat{s}}^s L^{u(\cdot)} f(X^{t,x}_{r,u(\cdot)})dr | \mathcal{F}^t_{\hat{s}}] \\ =f(X^{t,x}_{\hat{s},u(\cdot)}) &- \int_t^{\hat{s}} L^{u(\cdot)} f(X^{t,x}_{r,u(\cdot)})dr = M^f_{\hat{s}},\ P-a.s. \end{aligned}$$

That is, $\{M^f_s\}_{t\le s\le T}$ is a $\{\mathcal{F}^t_s\}$-martingale with mean $f(x)$. Thus we obtain Dynkin's formula

$$E[f(X^{t,x}_{s,u(\cdot)})] = f(x) + E[\int_t^s L^{u(\cdot)} f(X^{t,x}_{r,u(\cdot)})dr]. \tag{18}$$

Similarly, we can prove that for any function $f(t,x)$, which is bounded in x and right differentiable w.r.t. t, we have

$$\begin{aligned} &E[f(s, X^{t,x}_{s,u(\cdot)})] \\ =&f(t,x) + \int_t^s E[f_t(r, X^{t,x}_{r,u(\cdot)}) + L^{u(\cdot)} f(r, X^{t,x}_{r,u(\cdot)})]dr. \end{aligned} \tag{19}$$

Theorem 3: The value function $V(t,x)$ is right differentiable w.r.t. t, and is a classical solution of the following first-order nonlinear differential-difference equation

$$V^+_t(t,x) + \sup_{u\in U} L^u V(t,x) = 0,\ \forall(t,x) \in [0,T) \times B, \tag{20}$$

with boundary condition (4), where $V^+_t(\cdot,x)$ denotes the right derivative of V w.r.t. t.

Proof: Consider a sufficiently small $\delta > 0$, by (9) and for any $x \in B$, we have

$$\begin{aligned} &V(t+\delta,x) - V(t,x) \\ =&V(t+\delta,x) - \sup_{u(\cdot)\in\mathcal{U}_t} E\left[V\left(t+\delta, X^{t,x}_{t+\delta,u(\cdot)}\right)\right] + o(\delta). \end{aligned}$$

For any $u(\cdot) \in \mathcal{U}_t$ (with $u_t = u \in U$), we have

$$\begin{aligned} &\limsup_{\delta\to 0} \frac{1}{\delta}(V(t+\delta,x) - V(t,x)) \\ \le& \lim_{\delta\to 0} \frac{1}{\delta}\left(\sum_{x'\neq x} P\{X^{t,x}_{t+\delta,u(\cdot)} = x'\}(V(t+\delta,x) - V(t+\delta,x')) + o(\delta)\right) \\ =& \sum_{x'\neq x} \lambda_{xx'}(t,u)(V(t,x) - V(t,x')). \end{aligned}$$

From the definition of generator (17) of X_t and taking the supremum over $u \in U$, we have

$$\limsup_{\delta\to 0} \frac{1}{\delta}(V(t+\delta,x) - V(t,x)) + \sup_{u\in U} L^u V(t,x) \le 0. \tag{21}$$

Conversely, there exists $\hat{u}(\cdot) \in \mathcal{U}_t$ (with $\hat{u}_t = \hat{u} \in U$) such that

$$\begin{aligned} &\frac{1}{\delta}(V(t+\delta,x) - V(t,x)) \\ >&\frac{1}{\delta}\left[V(t+\delta,x) - E\left[V\left(t+\delta, X^{t,x}_{t+\delta,\hat{u}(\cdot)}\right)\right] - \varepsilon\delta + o(\delta)\right]. \end{aligned}$$

In the lower limit of $\delta \to 0$, we have

$$\liminf_{\delta\to 0} \frac{1}{\delta}(V(t+\delta,x) - V(t,x)) + L^{\hat{u}} V(t,x) > -\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary and taking the supremum over $u \in U$, we have

$$\liminf_{\delta\to 0} \frac{1}{\delta}(V(t+\delta,x) - V(t,x)) + \sup_{u\in U} L^u V(t,x) \ge 0. \tag{22}$$

Since the transition rates of X_t is continuous and $V \in [0,1]$, then

$$\begin{aligned} &|\sup_{u\in U} \sum_{x'\neq x} \lambda_{xx'}(t,u)(V(t,x) - V(t,x'))| \\ &\le \sup_{u\in U} \sum_{x'\neq x} \lambda_{xx'}(t,u) = \sup_{u\in U} |\lambda_{xx}(t,u)| < \infty. \end{aligned}$$

This, combining (21) and (22) leads to

$$\lim_{\delta\to 0} \frac{1}{\delta}(V(t+\delta,x) - V(t,x)) + \sup_{u\in U} L^u V(t,x) = 0. \tag{23}$$

Then, the limit $V^+_t(t,x) := \lim_{\delta\to 0} \frac{1}{\delta}(V(t+\delta,x) - V(t,x))$ exists and is finite and unique because of the uniqueness of supremum. Thus, the HJB equation (20) holds and the value function $V(t,x)$ is its classical solution. It is easy to see the boundary conditions are also satisfied. ■

Remark 6: In classical DP problems, the differentiabilities of the value function V w.r.t. the time variable and state variable are normally unattainable. One usually can only prove that the value function is a viscosity solution [8]. The existence and uniqueness of viscosity solutions to HJB equations can be found in [8] and some other literature, such as [18], [27] and [19]. In this article, we proved that the value function is a classical solution of HJB equation (20), and the uniqueness will be given in Theorem 5.

Since the transition rates of the process X_t is assumed to be stable, then the function $(u,t,x) \mapsto L^u V(t,x)$ is continuous w.r.t. u in U and measurable w.r.t. (t,x) in $[0,T) \times B$, so by the compactness assumption of U and the measurable selection theorem [2], there exists a measurable control function $\bar{u}(\cdot,\cdot)$ in $[0,T) \times B$ such that

$$L^{\bar{u}(t,x)} V(t,x) = \sup_{u \in U} L^u V(t,x), \quad \forall (t,x) \in [0,T) \times B. \tag{24}$$

Let

$$u^*(t,\omega) = \bar{u}(t, X_t(\omega)). \tag{25}$$

Then, $u^*(\cdot)$ is a deterministic Markov control, that is, $u^*(\cdot) \in \mathcal{M}$.

Next we will prove that the deterministic Markov control $u^*(\cdot)$ defined by (25) is an optimal control of Problem (S). This result provides the conjugacy of the optimal control in terms of the HJB equation and that of the utility surviving probability.

Theorem 4: The deterministic Markov control $u^*(\cdot) \in \mathcal{M}$ defined by (25) is the optimal control of Problem (S), i.e., such $u^*(\cdot)$ is the control such that $V(t,x) = J(t,x,u^*(\cdot)), \ \ \forall (t,x) \in [0,T) \times B$.

Proof: Let $u^*(\cdot)$ be the deterministic Markov control defined by (25). Then, by Dynkin's formula (19), we have

$$\begin{aligned}&E\left[V\left(T, X^{t,x}_{T,u^*(\cdot)}\right)\right]\\ =&V(t,x) + \int_t^T E\left[V_t^+\left(r, X^{t,x}_{r,u^*(\cdot)}\right) + L^{u^*(\cdot)} V\left(r, X^{t,x}_{r,u^*(\cdot)}\right)\right] dr.\end{aligned}$$

Note that this does not imply that $E[V(T, X^{t,x}_{T,u^*(\cdot)})] = V(t,x)$, since $u^*(\cdot)$ is the control such that $V_t^+(t,x) + L^{\bar{u}(t,x)}V(t,x) = 0$ holds only for $x \in B$.

To proceed, we consider, for a sufficiently small $\delta > 0$ such that $(T-t)/\delta$ is an integer, the discretization with step size of length δ,

$$\begin{aligned}&E\left[V\left(t+\delta, X^{t,x}_{t+\delta,u^*(\cdot)}\right)\right]\\ =&V(t,x) + [V_t^+(t,x) + L^{\bar{u}(t,x)}V(t,x)]\delta + o(\delta) = V(t,x) + o(\delta).\end{aligned}$$

Since $V(s,x') = 0$ for each $x' \in B^c$, we have

$$\sum_{x' \in B} V(t+\delta, x') P\{X^{t,x}_{t+\delta,u^*(\cdot)} = x'\} = V(t,x) + o(\delta).$$

Furthermore,

$$\begin{aligned}&\sum_{x' \in B} E[V(t+2\delta, X^{t+\delta,x'}_{t+2\delta,u^*(\cdot)})] P\{X^{t,x}_{t+\delta,u^*(\cdot)} = x'\}\\ =&E[V(t+\delta, X^{t,x}_{t+\delta,u^*(\cdot)})] + \sum_{x' \in B} P\{X^{t,x}_{t+\delta,u^*(\cdot)} = x'\}\cdot\\ &\qquad [V_t^+(t+\delta,x') + L^{\bar{u}(t+\delta,x')}V(t+\delta,x')]\delta + o(\delta)\\ =&V(t,x) + o(\delta).\end{aligned}$$

On the other hand,

$$\begin{aligned}&\sum_{i \in B} E[V(t+2\delta, X^{t+\delta,i}_{t+2\delta,u^*(\cdot)})] P\{X^{t,x}_{t+\delta,u^*(\cdot)} = i\}\\ =&\sum_{x' \in B} V(t+2\delta,x')(\sum_{i \in B} P\{X^{t,x}_{t+\delta,u^*(\cdot)} = i\} P\{X^{t+\delta,i}_{t+2\delta,u^*(\cdot)} = x'\})\\ =&\sum_{x' \in B} V(t+2\delta,x')(\sum_{i \in B} P\{X^{t+\delta,X_{t+\delta}}_{t+2\delta,u^*(\cdot)} = x', X^{t,x}_{t+\delta,u^*(\cdot)} = i\})\\ =&\sum_{x' \in B} V(t+2\delta,x') P\{X^{t+\delta,X_{t+\delta}}_{t+2\delta,u^*(\cdot)} = x', X^{t,x}_{t+\delta,u^*(\cdot)} \in B\}.\end{aligned}$$

Since for each $x \in B$, $V(T,x) = 1$, by iteration, we have

$$\begin{aligned}&V(t,x) + \frac{o(\delta)}{\delta}\\ =&P\{X^{T-\delta,X_{T-\delta}}_{T,u^*(\cdot)} \in B, \cdots, X^{t+\delta,X_{t+\delta}}_{t+2\delta,u^*(\cdot)} \in B, X^{t,x}_{t+\delta,u^*(\cdot)} \in B\}.\end{aligned}$$

In the limit of $\delta \to 0$, we have

$$V(t,x) = P\{X^{t,x}_{s,u^*(\cdot)} \in B, \forall s \in [t,T]\} = J(t,x,u^*(\cdot)). \tag{26}$$

Thus, $u^*(\cdot)$ is an optimal control of Problem (S). ■

Theorem 4 says that there exists an optimal deterministic Markov control of Problem (S); thus, the value function can be rewritten as

$$V(t,x) = \max_{u(\cdot) \in \mathcal{M}_t} J(t,x,u(\cdot)). \tag{27}$$

Theorem 3 says that V is a classical solution of HJB equation (20), and Theorem 4 gives the existence theorem of optimal control, based on which we have the following verification theorem.

Theorem 5: The value function V is the unique solution of HJB equation (20).

Proof: We need to prove that if there exists another bounded and measurable function $\psi(t,x)$ solving (20), which is right differentiable w.r.t. t for each $x \in B$, then

$$\psi(t,x) = V(t,x), \ \forall (t,x) \in [0,T] \times S. \tag{28}$$

In fact, we only need to prove that (28) holds for each $(t,x) \in [0,T) \times B$. By Dynkin's formula (19) and for each $u(\cdot) \in \mathcal{M}_t$, we have

$$\begin{aligned}&E[\psi(T, X^{t,x}_{T,u(\cdot)})]\\ =&\psi(t,x) + E[\int_t^T \psi_t^+(r, X^{t,x}_{r,u(\cdot)}) + L^{u(r,X_r)}\psi(r, X^{t,x}_{r,u(\cdot)})dr].\end{aligned}$$

Considering the discretization of the process $X^{t,x}_{s,u(\cdot)}$ with step size of length δ such that $(T-t)/\delta$ is an integer, we have

$$E[\psi(t+\delta, X^{t,x}_{t+\delta,u(\cdot)})] = \psi(t,x) + (\psi_t^+(t,x) + L^{u(t,x)}\psi(t,x))\delta + o(\delta),$$

since $\psi(s,x') = 0$ for each $x' \in B^c$, we have

$$\begin{aligned}&\sum_{x' \in B} \psi(t+\delta,x') P\{X^{t,x}_{t+\delta,u(\cdot)} = x'\}\\ =&\psi(t,x) + (\psi_t^+(t,x) + L^{u(t,x)}\psi(t,x))\delta + o(\delta).\end{aligned} \tag{29}$$

Since ψ solves (20), that is

$$\psi_t^+(t,x) + L^{u(t,x)}\psi(t,x) \leq 0,$$

then we have

$$\sum_{x' \in B} \psi(t+\delta,x') P\{X^{t,x}_{t+\delta,u(\cdot)} = x'\} \leq \psi(t,x) + o(\delta).$$

Similar as in the proof of Theorem 4, we have for each $u(\cdot)$

$$\psi(t,x) \geq P\{X^{t,x}_{s,u(\cdot)} \in B, \forall s \in [t,T]\}.$$

Taking the maximum over $u(\cdot) \in \mathcal{M}_t$, we have

$$\psi(t,x) \geq \max_{u(\cdot) \in \mathcal{M}_t} P\{X^{t,x}_{s,u(\cdot)} \in B, \forall s \in [t,T]\}. \tag{30}$$

Conversely, there exists a deterministic Markov control $u^*(\cdot) \in \mathcal{M}$, such that

$$\sup_{u \in U} L^u \psi(t,x) = L^{u^*(t,x)}\psi(t,x), \ \forall (t,x) \in [0,T) \times B.$$

Using the discretization method used to derive (26), we have

$$\psi(t,x) = P\{X^{t,x}_{s,u^*(\cdot)} \in B, \forall s \in [t,T]\}. \tag{31}$$

Combining (30) and (31), we have

$$\psi(t,x) = \max_{u(\cdot)\in\mathcal{M}_t} P\{X^{t,x}_{s,u(\cdot)} \in B, \forall s \in [t,T]\} = V(t,x).$$

That completes the proof. ■

It is not difficult to find that the u^* given in the Theorem 5 also is an optimal control of Problem (S).

IV. The case starting from outside of the target set

In the previous sections, we have discussed the stochastic optimal control problem whose utility functional is given by the probability on a given set B, if $X_0 = x_0 \in B$. However, x_0 does not necessarily belong to B. For the case that $x_0 \notin B$ and any $t_0 \in (0,T]$, we can consider the optimal control problem to find an optimal control to maximize the probability that $P\{X^{t_0,X_{t_0}}_{s,u(\cdot)} \in B, s \in [t_0,T]\}$.

In this section we will find the optimal control within $\mathcal{M}$. In fact

$$\begin{aligned}&P\{X^{t_0,X_{t_0}}_{s,u(\cdot)} \in B, s\in[t_0,T]\}\\ &=\sum_{x\in B} P\{X^{x_0}_{t_0,u(\cdot)} = x\}P\{X^{t_0,x}_{s,u(\cdot)} \in B, s\in[t_0,T]\}.\end{aligned} \tag{32}$$

Similar as in the proof of Theorem 1, we have

$$\begin{aligned}&\sup_{u(\cdot)\in\mathcal{M}}\left(\sum_{x\in B} P\{X^{x_0}_{t_0,u(\cdot)} = x\}P\{X^{t_0,x}_{s,u(\cdot)} \in B, s\in[t_0,T]\}\right)\\ &=\sup_{u(\cdot)\in\mathcal{M}}\left(\sum_{x\in B} P\{X^{x_0}_{t_0,u(\cdot)} = x\}\right.\\ &\qquad\left.\times\sup_{u(\cdot)\in\mathcal{M}_{t_0}} P\{X^{t_0,x}_{s,u(\cdot)} \in B, s\in[t_0,T]\}\right).\end{aligned}$$

We have proved that there exists $u^*(\cdot) \in \mathcal{M}_{t_0}$ such that

$$\sup_{u(\cdot)\in\mathcal{M}_{t_0}} P\{X^{t_0,x}_{s,u(\cdot)} \in B, s\in[t_0,T]\} = P\{X^{t_0,x}_{s,u^*(\cdot)} \in B, s\in[t_0,T]\}.$$

Denote the above probabilities by a_x; then, (32) turns out to be

$$\sum_{x\in B} a_x P\{X^{x_0}_{t_0,u(\cdot)} = x\}. \tag{33}$$

It turns out that the optimal control problem that we consider in this section is to maximize the utility functional: for $(t,x) \in [0,t_0]\times S$

$$\mathcal{J}(t,x,u(\cdot)) = E[\sum_{x'\in B} I_{\{X^{t,x}_{t_0,u(\cdot)}=x'\}}a_{x'}] =: E[g(X^{t,x}_{t_0,u(\cdot)})],$$

over $u(\cdot)\in\mathcal{M}_t$ and g is a bounded function. Then, the value function is

$$\mathcal{V}(t,x) = \sup_{u(\cdot)\in\mathcal{M}_t} \mathcal{J}(t,x,u(\cdot)). \tag{34}$$

With a similar proof to that of Theorems 1, 3, 4 and 5, we have the following theorem.

Theorem 6: The value function $\mathcal{V}(t,x)$ satisfies the DP equation

$$\mathcal{V}(t,x) = \sup_{u(\cdot)\in\mathcal{M}_t} E[\mathcal{V}(s, X^{t,x}_{s,u(\cdot)})],\ \forall s\in(t,t_0]. \tag{35}$$

The value function $\mathcal{V}$ is the unique solution of HJB equation

$$\begin{cases}\mathcal{V}^+_t(t,x) + \sup_{u\in U} L^u\mathcal{V}(t,x) = 0,\ (t,x)\in[0,t_0)\times S,\\ \mathcal{V}(t_0,x) = a_x,\ if\ x\in B\ or\ 0\ otherwise, t_0\in(0,T].\end{cases} \tag{36}$$

There exists a deterministic Markov control $u^*(\cdot)\in\mathcal{M}_t$ such that

$$\mathcal{V}(t,x) = \mathcal{J}(t,x,u^*(\cdot)),\quad (t,x)\in[0,t_0]\times S. \tag{37}$$

Remark 7: In fact, the discussion in this section covers the case that $t_0 = 0$ with $P\{X^{x_0}_{s,u(\cdot)} \in B, s\in[t_0,T]\} = 0$ satisfying the terminal condition in (36).

V. Examples

Example 1: Let $S = \{1,2,3\}$ be the state space of Markov chain X_t, $t\in[0,T]$ and $B = \{2\}$ as the target set. By the discussion given in the previous sections, the value function $V(t,x)$ satisfies the HJB equation (20). By the boundary condition of V, we have

$$V^+_t(t,2) + \sup_{u\in U} V(t,2)\lambda_{22}(t,u) = 0.$$

That is,

$$V(t,2) \geq \sup_{u(\cdot)\in\mathcal{M}_t} \exp\{\int_t^T \lambda_{22}(s,u(s))ds\}.$$

By the existence theorem of optimal control, we have

$$V(t,2) = \max_{u(\cdot)\in\mathcal{M}_t} \exp\{\int_t^T \lambda_{22}(s,u(s))ds\}.$$

Example 2: Consider a controlled time-homogeneous birth-and-death process, denoted by $\{X_t\}_{0\leq t\leq T}$, as an example of general controlled Markov processes considered in the main result of this article. The state space of the process X_t is $S = \{0,1,\cdots,K\}$, and the control domain is $U = [a,b]$. The transition rates are given by

$$\lambda_{xx'}(u) = \begin{cases} rx(u-\frac{2x}{K})^2, & x' = x+1,\\ dx(u-\frac{2x}{K})^2, & x' = x-1,\\ -(r+d)x(u-\frac{2x}{K})^2, & x' = x.\end{cases}$$

Our optimal control problem is to maximize the utility function

$$J(t,x,u(\cdot)) = P\{X^{t,x}_{s,u(\cdot)} \in B, \forall s\in[t,T]\},\ (t,x)\in[0,T]\times B,$$

over $u(\cdot)\in\mathcal{M}_t$. This example satisfies all conditions set in this article, and the corresponding value function is

$$V(t,x) = \max_{u(\cdot)\in\mathcal{M}_t} J(t,x,u(\cdot)).$$

Recall the DP equation (8) and HJB equation (20). By (24), the optimal control is the control maximizing

$$\begin{aligned}L^uV(t,x) =&\lambda_{x,x+1}(u)(V(t,x+1)-V(t,x))\\ &+\lambda_{x,x-1}(u)(V(t,x-1)-V(t,x))\\ =:&G_x(u-\frac{2x}{K})^2,\ x\in B,\end{aligned}$$

where $G_x = rx(V(t,x+1)-V(t,x)) + dx(V(t,x-1)-V(t,x))$. If G_x is positive, we need to find $u\in[a,b]$ to maximize $(u-\frac{2x}{K})^2$, and if G_x is negative, we need to find u to minimize $(u-\frac{2x}{K})^2$. Because of the right-continuity of X_t, it is easy to see from (25) and Theorem 4 that the control process is right continuous. Therefore, the control process is measurable. Unfortunately, the partial derivative in HJB equation is just a right-derivative, and the HJB equation is a backward differential-difference equation, so it cannot be used directly for the simulation of this stochastic optimal control problem. We can investigate the properties of optimal control and simulate value function using DP equation.

For a numerical experiment, consider $d = r = 0.04$, $K = 100$, $B = \{30,31,\cdots,60\}$, $T = 100$, $\Delta t = 0.01$, and $N = T/\Delta t$. In addition, let $a = 1$, $b = 2$, $\Delta u = 0.01$, and $U = \{1, 1.01, 1.02, \cdots, 2\}$. This model is discretized and naturally becomes a discrete-time MDP with a risk probability criterion. The one-step transition probabilities of $X_n = X_{n\Delta t}$ are given by

$$P_{xx'}(u) = \begin{cases} rx\Delta t(u - \frac{2x}{K})^2, & x' = x+1, \\ dx\Delta t(u - \frac{2x}{K})^2, & x' = x-1, \\ 1-(r+d)x\Delta t(u - \frac{2x}{K})^2, & x' = x. \end{cases}$$

The discrete-time version of DPP (8) (see Theorem 1) is stated in the following steps:
Step 1. Let $V(N,x) = 1$, for all $x \in B$;
Step 2. For $n \in \{N-1, \cdots, 1, 0\}$,

$$\begin{aligned} V(n,30) = \max_{u\in U}\{&0.04\cdot 30\cdot 0.1\cdot (u-0.6)^2\cdot V(n+1,31) \\ &+(1-0.08\cdot 30\cdot 0.1\cdot (u-0.6)^2)\cdot V(n+1,30)\}, \end{aligned}$$

$$\begin{aligned} V(n,60) = \max_{u\in U}\{&0.04\cdot 60\cdot 0.1\cdot (u-1.2)^2\cdot V(n+1,59) \\ &+(1-0.08\cdot 60\cdot 0.1\cdot (u-1.2)^2)\cdot V(n+1,60)\}, \end{aligned}$$

for $x \in \{31, \cdots, 59\}$,

$$\begin{aligned} V(n,x) = \max_{u\in U}\{&0.04x\cdot 0.1\cdot (u-\frac{x}{50})^2\cdot V(n+1,x+1) \\ &+0.04x\cdot 0.1\cdot (u-\frac{x}{50})^2\cdot V(n+1,x-1) \\ &+(1-0.08x\cdot 0.1\cdot (u-\frac{x}{50})^2)\cdot V(n+1,x)\}, \end{aligned}$$

record each optimal u;
Step 3. Plot $V(0,x)$.
We calculate $V(0,x)$ for $U = \{1, 1.01, 1.02, \cdots, 2\}$ and plot the graph in Fig. 1 as the curves in red dotted line. We also take u =1, 1.5 and 2 as three different fixed values and calculate $V(0,x)$ according to the above algorithm without seeking optimal control u and plot the graph $x \mapsto V(0,x)$ in Fig. 1 as the curves in orange, blue and black dotted lines.

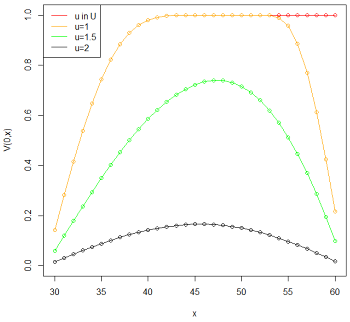

Fig. 1. V(0,x) for different control domain U.

Remark 8: 1). The difference between the optimal surviving probabilities with a proper control (taking $U = \{1, 1.01, \cdots, 2\}$) and the surviving probabilities have a fixed u is clearly shown especially for x being near to the top of the target set. It is easy to see that the red line is higher than and equal to other three lines for each x.
2). The optimal control process is recorded as a matrix, which can provide the optimal policy for each DP problem initiated by each state and time.

We also calculate the optimal control problems of surviving probabilities with $U = \{1, 1.01, \cdots, 2\}$, $N = \{30, 31, \cdots, 60\}$ and different values of d (r is assumed to be the same as before): $d = 0.6r, 0.8r, r, 1.2r, 1.4r$, respectively. We plot them in Fig.2.

It is noted that the probability $V(0,x)$ is increasing with the decrease of d for each $x \in B$, which means decreasing the death rates is benefit to the controlled Markov chain staying in a safety range under the condition that the birth rate remains unchanged.

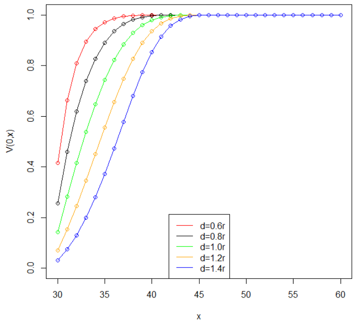

Fig. 2. V(0,x) for different death rates d.

We also simulate value functions according to the above algorithm (Steps 1-3) for $B = \{0, 1, \cdots, 30\}$ and calculate them with $U = \{1, 1.01, \cdots, 2\}$ and different values of d (r is assumed to be the same as before): $d = 0.6r, 0.8r, r, 1.2r, 1.4r$, respectively. We plot them in Fig.3.

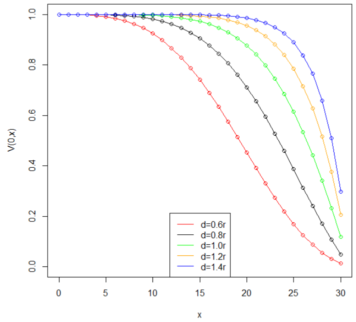

Fig. 3. V(0,x) for different death rates d.

Different from Fig.2, the value function $V(0,x)$ in Fig.3 is increasing with the increase of d for each $x \in B$ under the condition that r remains unchanged. An example of the model in Fig.2 is the human population model staying in a set of a relatively moderate or large size. In this case, decreasing the death rate can make a substantial difference only when the population has a relatively moderate size. While the model in Fig.3 show, e.g., in the cancer cell population model, it is desirable to have less number or extinction of cancer cells. In the latter case, increasing death rate improves the probability of keeping the size of cancer cells to be in the safety target set.

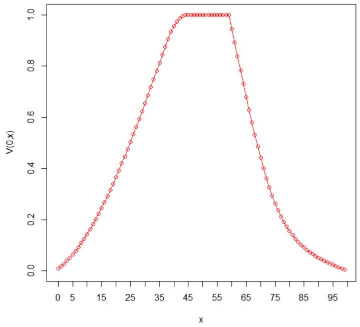

Fig. 4. Graph of $\boldsymbol{\mathcal{V}(0,x)}$, $\boldsymbol{x \in \{0,1,\cdots,100\}}$.

We also consider the case starting from the outside of the target set B and use the value function $V(0,x)$ with $U = \{1, 1.01, 1.02, \cdots, 2\}$ (see the red line in Fig.1) as the terminal condition of the value function $\mathcal{V}(T,x)$. Using the algorithm as explained in Step 2 with $U = \{1, 1.01, \cdots, 2\}$, we carry out our numerical computation and plot our result in Fig. 4. Note that, in this case, the algorithm focus on the state space S, not just on the target set B.

VI. Conclusion and further considerations

In this article, we consider continuous-time MDPs and derive the DPP and HJB equation for optimal surviving probability $V(t,x) = \sup_{u(\cdot)\in\mathcal{U}_t} P\{X^{t,x}_{s,u(\cdot)} \in B, \forall s \in [t,T]\}$ instead of using the optimality equations as in [15] and [16]. The optimality criterion we consider is the risk probability that the first exit time from a target set of the controlled Markov chains exceeds a fixed value. Such a setup is applicable in many real-world contexts. The main effort of this article is to establish the DPP, derive the HJB equation, prove the existence of optimal controls, and verify the value function is the unique solution of the HJB equation. The problem is not covered by the traditional stochastic optimal problem and associated DPP and HJB equation. It is also not covered by the risk probability considered by Huo *et al.* [15], [16]. In fact, as $V(t,x)$ depends on the surviving set B, we can define a set function as $V_{t,x}(B) = V(t,x)$. Then it is easy to see that $V_{t,x} : \mathcal{V}(S) \to [0,1]$ is a superadditive capacity. In this sense, we obtained the HJB equation for a superadditive capacity given by an optimal surviving probability. This is in contrast to the traditional stochastic optimal control problems, where the value function is a sublinear expectation operator [23]. Given the recent progress of the ergodic theory of sublinear semigroup and capacity [10], [11], and that the superlinear semigroup and the sublinear semigroup are conjugate to each other, it would be very interesting to ask whether or not there exists an invariant superlinear distribution μ such that $V_{t,\mu} = \mu$. Here $V_{t,\mu}(B) = (\mu V_{t,\cdot})(B)$. If so, is μ a continuous distribution and is it ergodic? The existence of such an invariant distribution is important in applications giving the equilibrium of a controlled superlinear process. We will publish this result in a further publication. The DPP and the HJB equations that we established in this article will be important tools for the analysis of invariant distribution and its ergodicity.

In this article, we mainly study the stochastic control problem of a risk probability criterion of a given controlled Markov process model. To link our results directly with applications, we should estimate the transition rates by studying some inverse problems and carrying out statistical analysis of datasets. This is clearly very important and worth pursuing in a future project. Controllability and observability for stochastic control systems are also interesting problems to investigate; we refer to [20] and [32] and will expand this aspect in the future.

Acknowledgment

The authors would like to acknowledge the anonymous referees for their constructive comments, which led to substantial improvements of this article.

References

[1] M. Abu Alsheikh, D. Hoang, D. Niyato, H. Tan, and S. Lin. Markov decision processes with applications in wireless sensor networks: A survey. *IEEE Communications Surveys Tutorials*, 17:1239–1267, 2015.

[2] B. Barmish. Measurable selection theorems and their application to problems of guaranteed performance. *IEEE Transactions on Automatic Control*, 23:685–687, 1978.

[3] R. Bellman. On the theory of dynamic programming. *Proceedings of the National Academy of Sciences*, 38:716–719, 1952.

[4] A. Bhabak and S. Saha. Continuous-time zero-sum games with probability criterion. *Stochastic Analysis and Applications*, 39:1130–1143, 2021.

[5] M. Bouakiz and Y. Kebir. Target-level criterion in Markov decision processes. *Journal of Optimization Theory and Applications*, 86:1–15, 1995.

[6] N. Bäuerle and U. Rieder. *Markov Decision Processes with Applications to Finance*. Springer Berlin Heidelberg, 2017.

[7] R. Cogill and C. Peng. Reversible Markov decision processes with an average-reward criterion. *SIAM Journal on Control and Optimization*, 51:402–418, 2013.

[8] M. Crandall, H. Ishii, and P. Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull.amer.math.soc*, 27:1–67, 1992.

[9] M. El Chamic, Y. Yu, B. Açıkmeşe, and M. Ono. Controlled Markov processes with safety state constraints. *IEEE Transactions on Automatic Control*, 64:1003–1018, 2019.

[10] C. Feng, P. Wu, and H. Zhao. Ergodicity of invariant capacities. *Stochastic Processes and their Applications*, 130:5037–5059, 2020.

[11] C. Feng and H. Zhao. Ergodicity of sublinear Markovian semigroups. *SIAM Journal on Mathematical Analysis*, 53:5646–5681, 2021.

[12] W. Fleming and H. Soner. *Controlled Markov processes and viscosity solutions*, volume 25. Springer Science & Business Media, 2006.

[13] X. Huang and X. Guo. Nonzero-sum stochastic games with probability criteria. *Dynamic Games and Applications*, 10:509–527, 2020.

[14] Y. Huang and X. Guo. Optimal risk probability for first passage models in semi-Markov decision processes. *Journal of Mathematical Analysis and Applications*, 359:404–420, 2009.

[15] H. Huo and X. Guo. Risk probability minimization problems for continuous-time Markov decision processes on finite horizon. *IEEE Transactions on Automatic Control*, 65:3199–3206, 2020.

[16] H. Huo, X. Zou, and X. Guo. The risk probability criterion for discounted continuous-time Markov decision processes. *Discrete Event Dynamic Systems*, 27:675–699, 2017.

[17] A. Jaśkiewicz and A. Nowak. Constrained Markov decision processes with expected total reward criteria. *SIAM Journal on Control and Optimization*, 57:3118–3136, 2019.

[18] S. Lv. Two-player zero-sum stochastic differential games with regime switching. *Automatica*, 114:108819, 2020.

[19] S. Lv, Z. Wu, and Q. Zhang. Optimal switching under a hybrid diffusion model and applications to stock trading. *Automatica*, 94:361–372, 2018.

[20] Q. Lü and X. Zhang. *Mathematical Control Theory for Stochastic Partial Differential Equations*. 2021.

[21] B. Miller. Optimization of queuing system via stochastic control. *Automatica*, 45:1423–1430, 2009.

[22] B. Miller, G. Miller, and K. Siemenikhin. Towards the optimal control of Markov chains with constraints. *Automatica*, 46:1495–1502, 2010.

[23] S. Peng. *Nonlinear Expectations and Stochastic Calculus under Uncertainty: with Robust CLT and G-Brownian Motion*. Springer, 2019.

[24] A. Piunovskiy. Multicriteria impulsive control of jump Markov processes. *Mathematical Methods of Operations Research*, 60:125–144, 2004.

[25] A. Piunovskiy and Y. Zhang. Discounted continuous-time Markov decision processes with unbounded rates: The convex analytic approach. *SIAM Journal on Control and Optimization*, 49:2032–2061, 2011.

[26] Y. Shen, W. Stannat, and K. Obermayer. Risk-sensitive Markov control processes. *SIAM Journal on Control and Optimization*, 51:3652–3672, 2013.

[27] Q. Song, R. Stockbridge, and C. Zhu. On optimal harvesting problems in random environments. *SIAM journal on control and optimization*, 49:859–889, 2011.

[28] X. Wen, X. Guo, and L. Xia. Optimal dividend problems with a risk probability criterion. *Naval Research Logistics*, 69:421–430, 2022.

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

[30] G. Yin and Q. Zhang. *Continuous-Time Markov Chains and Applications: a Two-Time-Scale Approach*, volume 37. Springer, 2013.

[31] J. Yong and X. Zhou. *Stochastic Controls: Hamiltonian Systems and HJB Equations*, volume 43. Springer Science & Business Media, 1999.

[32] R. Zhang and L. Guo. Controllability of stochastic game-based control systems. *SIAM Journal on Control and Optimization*, 57:3799–3826, 2019.