Stochastic Processes and their Applications 91 (2001) 1–20

stochastic processes and their applications

www.elsevier.com/locate/spa

Two-sided taboo limits for Markov processes and associated perfect simulation

Peter W. Glynn[a,*,1], Hermann Thorisson[b,2]

[a] Department of Management Science and Engineering, Terman Engineering Center, Stanford University, Room 429A, Stanford, CA 94305-4023, USA

[b] Science Institute, University of Iceland, Dunhaga 3, 107 Reykjavik, Iceland

Received 1 February 1999; received in revised form 16 May 2000; accepted 19 May 2000

Abstract

In this paper, we study the two-sided taboo limit processes that arise when a Markov chain or process is conditioned on staying in some set A for a long period of time. The taboo limit is time-homogeneous after time 0 and time-inhomogeneous before time 0. The time-reversed limit has this same qualitative structure. The precise transition structure at the taboo limit is identified in the context of discrete- and continuous-time Markov chains, as well as diffusions. In addition, we present a perfect simulation algorithm for generating exact samples from the quasi-stationary distribution of a finite-state Markov chain.

Keywords: Markov chains; Markov processes; Quasi-stationary distribution; Eigenvalues; Perron–Frobenius theory; Perfect simulation; Diffusions

1. Introduction

In many applications settings, it is of interest to approximate the behavior of a process conditioned on its staying out of some specific subset of the state space. For example, one might model the fish population in a lake as a stochastic process in which eventual extinction is a certainty. To properly understand the dynamics of an existing such population, it may be appropriate to study the population process conditioned on its not having gone extinct over some long period of time.

This topic has attracted significant attention within the Markov chain literature. In particular, if one conditions a Markov chain on staying out of a subset, the resulting (marginal) distribution of the chain can often be approximated by what is known as the “quasi-stationary distribution”.

* Corresponding author. Tel.: +1-650-725-0554; fax: +1-650-723-4107.

E-mail address: glynn@leland.stanford.edu (P.W. Glynn).

[1] This research was supported by the U.S. Army Research Office under contract No. DAAGSS-97-1-0377 and by the National Science Foundation under Grant No. PMS-9704732.

[2] www.hi.is/ ~ hermann

PII: S0304-4149(00)00050-8

The question of existence of quasi-stationary distributions can be settled by appealing to the theory at R-recurrence for general non-negative kernels; see, for example, Seneta and Vere-Jones (1966), Tweedie (1974), Nummelin and Arjas (1976), Nummelin and Tweedie (1978), and Nummelin (1984).

In this paper, our emphasis is on studying not just the marginal distribution but rather the entire joint distribution of the process, conditioned on staying out of some specific set A^c. Specifically, suppose we have conditioned the process to remain in A over the entire interval $[0,t]$ (with t large). We develop an approximation for the behavior of the process both prior to t and subsequent to t. (The quasi-stationary distribution approximates the behavior at t.) The "two-sided taboo limit" that arises is then an approximation to the time-dependent behavior of the original tabooed process.

We establish existence of such two-sided taboo limits in the Markov context. We show that the two-sided taboo limit has time-homogeneous transition probabilities to the right (in the region in which the taboo limit is approximating the post-t behavior of the original process) and time-inhomogeneous transition probabilities from the left (in the region in which the taboo limit is approximating the pre-t behavior at the original process). In fact, we are able to explicitly compute the above transition structure for discrete- and continuous-time Markov chains, as well as diffusions; see Theorems 2, 6, and 7. In addition, we show in the Markov chain context that the same qualitative structure is inherited by the time-reversed process, and explicitly compute its transition structure; see Theorems 2 and 6.

In the Markov chain context, we are also able to develop a corresponding two-sided taboo limit process that arises as an approximation to the original process when conditioned on exiting to A^c at time t (for t large); see Theorem 4 and Proposition 3. In addition, we show that the two taboo limits are related through a "geometric time-shift" in the discrete time Markov chain setting (Theorem 5). This is a particular case of a more general phenomenon that is studied in greater detail in our companion paper; see Glynn and Thorisson (2000).

Finally, we develop an exact sampling (or perfect simulation) algorithm for producing observations from the quasi-stationary distribution (and, more generally, from the two-sided taboo limit process) in the setting in which A is finite, irreducible and aperiodic; see Section 5. Perhaps surprisingly, the algorithm requires no apriori knowledge of the solution of the eigenvalue problem that characterizes the quasi-stationary distribution. This algorithm is a complement to the recently developed Propp and Wilson (1996) algorithm for exact sampling from the stationary distribution of a Markov chain. As in the Propp–Wilson context, the algorithm relies on a "backwards coupling" or "coupling from-the-past".

In our companion paper (Glynn and Thorisson, 2000), many of these issues are considered in the regenerative setting. In contrast to the theory developed in this paper, the results there are most naturally expressed on the time scale of regenerative cycles. That paper also considers the concept of "taboo-stationarity" for general processes. For a coupling approach to taboo limits, see Thorisson (2000).

This paper is organized as follows. In Section 2, we review the basic concepts from the theory of quasi-stationarity that we will need. Section 3 develops the relevant two-sided taboo limit theory in the general state space discrete-time Markov chain setting. Much of the theory is extended to continuous-time Markov chains and diffusions in Section 4. Finally, Section 5 presents the perfect simulation algorithm for discrete-time Markov chains with A finite, irreducible and aperiodic.

2. Quasi-stationary distributions for discrete-time Markov chains

Let $X = (X_n\colon n \geqslant 0)$ be a (time-homogeneous) discrete-time Markov chain taking values in a complete separable metric space S. For (Borel measurable) $A \subseteq S$, let

$$\Gamma = \inf\{n \geqslant 0\colon X_n \in A^c\}$$

be the first exit time of X from the set A. To develop approximations for the process X conditional on $\Gamma > n$, it seems clear that a key point is understanding the tail behavior of the r.v. Γ.

To this end, consider the sub-stochastic kernel $K = (K(x, \mathrm{d}y)\colon x, y \in A)$, obtained by restricting P to A, where $P = (P(x, \mathrm{d}y)\colon x, y \in S)$ is the (one-step) transition kernel of X. Specifically,

$$K(x, \mathrm{d}y) = P(x, \mathrm{d}y) \quad \text{for } x, y \in A.$$

To analyze K, we make an assumption that essentially requires that we be able to solve a certain eigenvalue problem:

A1. There exists $\lambda \in (0, 1]$, a strictly positive function $h = (h(x)\colon x \in A)$, and a non-trivial non-negative measure $v = (v(\mathrm{d}x)\colon x \in A)$ such that

$$\int_A K(x, \mathrm{d}y)h(y) = \lambda h(x) \quad \text{for } x \in A,$$

$$\int_A v(\mathrm{d}x)K(x, \mathrm{d}y) = \lambda v(\mathrm{d}y) \quad \text{for } y \in A.$$

Under A1, it is easily seen that $G = (G(x, \mathrm{d}y)\colon x, y \in A)$ defined by

$$G(x, \mathrm{d}y) = \lambda^{-1} K(x, \mathrm{d}y)h(y)/h(x) \quad \text{for } x, y \in A$$

is a stochastic kernel on A possessing an invariant measure $\eta = (\eta(\mathrm{d}x)\colon x \in A)$ given by

$$\eta(\mathrm{d}x) = v(\mathrm{d}x)h(x) \quad \text{for } x \in A.$$

Let $\boldsymbol{P}_x(\cdot)$ be the probability measure on the path-space of X (that is, on S^∞) under which X evolves as a Markov chain with transition kernel P, conditional on $X_0 = x \in S$. Similarly, let $\tilde{\boldsymbol{P}}_x(\cdot)$ be the probability on the path-space of X under which X is a Markov chain with transition kernel G, conditional on $X_0 = x \in A$. Note that

$$\begin{aligned} &\boldsymbol{P}_x(X_1 \in \mathrm{d}x_1, \ldots, X_n \in \mathrm{d}x_n, \Gamma > n) \\ &\quad = K(x, \mathrm{d}x_1)K(x_1, \mathrm{d}x_2) \ldots K(x_{n-1}, \mathrm{d}x_n) \\ &\quad = \lambda^n h(x) G(x, \mathrm{d}x_1) \ldots G(x_{n-1}, \mathrm{d}x_n)/h(x_n) \\ &\quad = \lambda^n h(x) \tilde{\boldsymbol{E}}_x[I(X_1 \in \mathrm{d}x_1, \ldots, X_n \in \mathrm{d}x_n)/h(X_n)] \end{aligned} \tag{2.1}$$

for $x, x_1, \ldots, x_n \in A$ where $\tilde{\boldsymbol{E}}_x[\,\cdot\,]$ is the expectation operator associated with $\tilde{\boldsymbol{P}}_x(\cdot)$. In particular, the above identity implies that for $x \in A$,

$$\boldsymbol{P}_x(\Gamma > n) = \lambda^n h(x) \tilde{\boldsymbol{E}}_x[1/h(X_n)].$$

To obtain precise tail asymptotics for the r.v. Γ, we require the following additional assumptions:

A2. X is an aperiodic positive recurrent Harris chain on A under G, and therefore η can be assumed to be normalized so that η is a probability.
A3. The measure v is finite, and can thus be assumed to be a probability.

Thus, under A2 (see Meyn and Tweedie (1993) for basic properties of Harris chains), for v a.e. x

$$\tilde{\boldsymbol{E}}_x[1/h(X_n)] \to \int_A \eta(\mathrm{d}x)/h(x) = 1 \quad \text{as } n \to \infty.$$

We therefore arrive at the conclusion that for v a.e. x,

$$\boldsymbol{P}_x(\Gamma > n) \sim \lambda^n\, h(x) \quad \text{as } n \to \infty. \tag{2.2}$$

The quasi-stationary distribution of X with respect to A is an approximation to the conditional marginal distribution $\boldsymbol{P}_x(X_n \in \cdot \,|\, \Gamma > n)$. In particular, it follows from (2.1) that

$$\boldsymbol{P}_x(X_n \in B | \Gamma > n) = \frac{\tilde{\boldsymbol{E}}_x[I(X_n \in B)/h(X_n)]}{\tilde{\boldsymbol{E}}_x[1/h(X_n)]} \quad \text{for } B \subseteq S.$$

Applying A2 and A3, it is evident that for v a.e. x

$$\boldsymbol{P}_x(X_n \in \cdot\,|\Gamma > n) \xrightarrow{\text{t.v.}} v(\cdot) \quad \text{as } n \to \infty$$

where $\xrightarrow{\text{t.v.}}$ denotes convergence in the sense of total variation. It is easy to verify that for $n \geqslant 1$,

$$\boldsymbol{P}_v(X_n \in \cdot\,|\Gamma > n) = v(\cdot), \tag{2.3}$$

where $\boldsymbol{P}_v(\cdot)$ is the probability on the path-space of X under which X has transition kernel P and initial distribution v.

Definition 1. A non-trivial measure v satisfying (2.3) is called a *quasi-stationary distribution* of X (with respect to A).

We summarize the discussion thus far with the following proposition.

Proposition 1. *Assume* A1–A3 *are satisfied. Then, there exists* $\alpha \geqslant 0$ *such that for* v *a.e.* x,

(i) $\boldsymbol{P}_x(\Gamma > n) \sim \mathrm{e}^{-\alpha n} h(x)$ *as* $n \to \infty$;
(ii) $\boldsymbol{P}_x(X_n \in \cdot\,|\Gamma > n) \xrightarrow{\text{t.v.}} v(\cdot)$ *as* $n \to \infty$.

Thus, under A1–A3, the quasi-stationary distribution v also appears as a limit distribution for the “tabooed chain”.

Remark 1. Versions of Proposition 1 are well known in the literature; see, for example, Section 6.7 of Nummelin (1984).

Remark 2. Assumptions A1–A3 are automatic when K is a finite irreducible aperiodic matrix. In this setting, these assumptions are an immediate consequence of the Perron–Frobenius theorem for finite matrices; see the appropriate appendix of Karlin and Taylor (1975) for details.

Remark 3. Sufficient conditions for A1–A3 in the countable state space and general state-space setting can be found in Section 2 of Ney and Nummelin (1987); see also Ferrari et al. (1996).

3. Two-sided taboo limits for discrete-time Markov chains

In constructing an approximation to the taboo behavior of X, conditioned on $\Gamma > n$, it is natural to use a two-sided limit process in which both the behavior of X prior to n and subsequent n can be jointly considered.

For $x \in A$, let $\tilde{\boldsymbol{P}}_{x,n}(\cdot)$ be the probability on the path-space of X, conditioned on $X_0 = x$, under which X evolves as a time-inhomogeneous Markov chain, using the transition kernel G up to and including the nth transition, and subsequently the transition kernel P. Specifically, for $x, x_1, \ldots, x_n \in A$ and $x_{n+1}, \ldots, x_{n+m} \in S$

$$\tilde{\boldsymbol{P}}_{x,n}(X_1 \in \mathrm{d}x_1, \ldots, X_n \in \mathrm{d}x_n, \ldots, X_{n+m} \in \mathrm{d}x_{n+m})$$

$$= G(x, \mathrm{d}x_1) \ldots G(x_{n-1}, \mathrm{d}x_n) P(x_n, \mathrm{d}x_{n+1}) \ldots P(x_{n+m-1}, \mathrm{d}x_{n+m}).$$

Relation (2.1) implies that

$$\boldsymbol{P}_x(X \in \cdot \,|\Gamma > n) = \frac{\tilde{\boldsymbol{E}}_{x,n}[I(X \in \cdot)/h(X_n)]}{\tilde{\boldsymbol{E}}_{x,n}[1/h(X_n)]} \quad \text{for } x \in A. \tag{3.1}$$

As indicated above, we will use a two-sided limit process to approximate $\boldsymbol{P}_x((X_{n+m}: -k \leqslant m < \infty) \in \cdot |\Gamma > n)$ for each k when n is large. Specifically, in a (slight) abuse of notation, let $X = (X_n: -\infty < n < \infty)$ be a two-sided stochastic sequence. Assume A1–A3 and suppose that $\tilde{\boldsymbol{P}}$ is the probability on the path-space of the two-sided process X under which

$$\tilde{\boldsymbol{P}}(X_{-k} \in \mathrm{d}x_{-k}, \ldots, X_k \in \mathrm{d}x_k)$$

$$= \eta(\mathrm{d}x_{-k}) G(x_{-k}, \mathrm{d}x_{-k+1}) \ldots G(x_{-1}, \mathrm{d}x_0) P(x_0, \mathrm{d}x_1)$$

$$\ldots P(x_{k-1}, \mathrm{d}x_k) \quad \text{for } x_{-k}, \ldots, x_0 \in A \text{ and } x_1, \ldots, x_k \in S.$$

Let $\boldsymbol{P}^*$ be the probability on the path-space of X defined by

$$\boldsymbol{P}^*(X \in \cdot) = \frac{\tilde{\boldsymbol{E}}[I(X \in \cdot)/h(X_0)]}{\tilde{\boldsymbol{E}}[1/h(X_0)]},$$

where $\tilde{\boldsymbol{E}}(\cdot)$ is the expectation operator corresponding to $\tilde{\boldsymbol{P}}$.

Theorem 1. *Fix* $k \geqslant 0$. *Under* A1–A3, *for* ν *a.e.* x,

$$\boldsymbol{P}_x((X_{n-k},\ldots,X_{n+k})\in\cdot\,|\Gamma > n) \stackrel{\text{t.v.}}{\to} \boldsymbol{P}^*((X_{-k},\ldots,X_k)\in\cdot)$$

as $n \to \infty$.

Proof. The theorem is an easy consequence of (3.1). Specifically, for f bounded (and measurable),

$$\boldsymbol{E}_x[f(X_{n-k},\ldots,X_{n+k})|\Gamma > n] = \frac{\tilde{\boldsymbol{E}}_{x,n}[f(X_{n-k},\ldots,X_{n+k})/h(X_n)]}{\tilde{\boldsymbol{E}}_{x,n}[1/h(X_n)]}.$$

But

$$\tilde{\boldsymbol{E}}_{x,n}[f(X_{n-k},\ldots,X_{n+k})/h(X_n)] = \tilde{\boldsymbol{E}}_x[g(X_{n-k})],$$

where

$$g(x) = \boldsymbol{E}^*[f(X_{-k},\ldots,X_k)/h(X_0)|X_{-k} = x].$$

Since, for ν a.e. x, $\tilde{\boldsymbol{E}}_x[1/h(X_n)] \to \boldsymbol{E}^*[1/h(X_0)]$ and

$$\tilde{\boldsymbol{E}}_x[g(X_{n-k})] \to \int_A \nu(\mathrm{d}x)g(x) = \boldsymbol{E}^*[g(X_{-k})] = \boldsymbol{E}^*[f(X_{-k},\ldots,X_k)]$$

as $n \to \infty$ uniformly in f bounded in absolute value by one, the theorem follows. □

In order to develop additional insight into the probability $\boldsymbol{P}^*$, we present the following result. In preparation for the statement of the theorem, let $Y = (Y_n\colon -\infty < n < \infty)$ be the two-sided time-reversed process corresponding to X, given by $Y_n = X_{-n}$ for $n \in \mathbb{Z}$. Because S is Polish, we can assert existence of the regular conditional distributions

$$R_k(x, \mathrm{d}y) = \boldsymbol{P}_\eta(X_k \in \mathrm{d}y|X_{k+1} = x),$$

$$R(x, \mathrm{d}y) = \tilde{\boldsymbol{P}}(X_{-1} \in \mathrm{d}y|X_0 = x)$$

for $x, y \in S$ and $k \geqslant 0$. (Here, $\boldsymbol{P}_\eta(\cdot)$ is the probability under which X is initiated with distribution η and makes transitions according to the kernel P.) Also, for $k \in \mathbb{Z}$, put

$$w_k(x) = \tilde{\boldsymbol{E}}[1/h(X_0)|X_k = x].$$

Theorem 2. *Assume* A1–A3. *Then*,

(i) $\boldsymbol{P}^*(X_{k+1}\in\mathrm{d}y|X_j\colon j\leqslant k) = P(X_k, \mathrm{d}y)$ *a.s. for* $k\geqslant 0$;
(ii) $\boldsymbol{P}^*(X_{k+1}\in\mathrm{d}y|X_j\colon j\leqslant k) = G(X_k, \mathrm{d}y)w_{k+1}(y)/w_k(X_k)$ *a.s. for* $k \leqslant -1$;
(iii) $\boldsymbol{P}^*(Y_{k+1}\in\mathrm{d}y|Y_j\colon j\leqslant k) = R(Y_k, \mathrm{d}y)$ *a.s. for* $k\geqslant 0$;
(iv) $\boldsymbol{P}^*(Y_{k+1}\in\mathrm{d}y|Y_j\colon j\leqslant k) = R_{-k-1}(Y_k, \mathrm{d}y)w_{-k-1}(y)/w_{-k}(Y_k)$ *a.s. for* $k \leqslant -1$.

Proof. For $\mathscr{F}$ a sub-σ-field of $\sigma(X_j\colon -\infty < j < \infty)$, it is well known that

$$\boldsymbol{P}^*(X_k \in \mathrm{d}y|\mathscr{F}) = \frac{\tilde{\boldsymbol{E}}[I(X_k \in \mathrm{d}y)/h(X_0)|\mathscr{F}]}{\tilde{\boldsymbol{E}}[1/h(X_0)|\mathscr{F}]} \quad \text{a.s.;}$$

see, for example, p. 171 of Brémaud (1980). It follows that for $k \geqslant 0$,

$$\begin{aligned}&\boldsymbol{P}^*(X_{k+1} \in \mathrm{d}y | X_j \colon j \leqslant k)\\&\quad = \frac{1/h(X_0)\tilde{\boldsymbol{P}}(X_{k+1} \in \mathrm{d}y | X_j \colon j \leqslant k)}{1/h(X_0)}\\&\quad = \tilde{\boldsymbol{P}}(X_{k+1} \in \mathrm{d}y | X_k) = P(X_k, \mathrm{d}y) \quad \text{a.s.},\end{aligned}$$

proving (i); (iii) follows similarly. For (ii), note that for $k \leqslant -1$,

$$\begin{aligned}&\boldsymbol{P}^*(X_{k+1} \in \mathrm{d}y | X_j \colon j \leqslant k)\\&\quad = \frac{\tilde{\boldsymbol{E}}[I(X_{k+1} \in \mathrm{d}y)\tilde{\boldsymbol{E}}[1/h(X_0) | X_j \colon j \leqslant k+1] | X_j \colon j \leqslant k]}{\tilde{\boldsymbol{E}}[1/h(X_0) | X_j \colon j \leqslant k]}\\&\quad = \frac{\tilde{\boldsymbol{E}}[I(X_{k+1} \in \mathrm{d}y) w_{k+1}(y) | X_j \colon j \leqslant k]}{w_k(X_k)}\\&\quad = G(X_k, \mathrm{d}y)\,\frac{w_{k+1}(y)}{w_k(X_k)} \quad \text{a.s.};\end{aligned}$$

(iv) follows analogously. □

According to Theorem 2, the taboo process X can be approximated, in the limit, by a Markov chain for which the forward evolution is time- inhomogeneous before time 0 and time-homogeneous after time 0. The time reversed chain Y has exactly the same qualitative structure. In addition,

$$w_k(y) \to \tilde{\boldsymbol{E}}[1/h(X_0)] \quad \text{for} \quad \eta \text{ a.e. } y \text{ as } k \to \infty,$$

see p. 111 of Nummelin (1984). Furthermore, $(\tilde{\boldsymbol{E}}[1/h(X_0) | X_j \colon j \leqslant k] \colon k \leqslant 0)$ is a backwards martingale, so that

$$\begin{aligned}w_k(X_k) &= \tilde{\boldsymbol{E}}[1/h(X_0) | X_j \colon j \leqslant k]\\&\to \tilde{\boldsymbol{E}}[1/h(X_0) | \mathscr{F}_{-\infty}] \quad \text{a.s. as } k \to -\infty;\end{aligned}$$

see, for example, p. 340 of Chung (1974). (Here, $\mathscr{F}_{-\infty}$ is the largest σ-field contained in $\sigma(X_j \colon j \leqslant k)$ for all $k \leqslant 0$.) Because of A2, $\mathscr{F}_{-\infty}$ is trivial, so for η a.e. y,

$$\frac{w_k(X_k)}{w_k(y)} \to 1 \quad \text{a.s. as } k \to -\infty.$$

Thus, while the limit is time-inhomogeneous before time 0, it is "asymptotically time-homogeneous to the left". A similar conclusion holds for the reversed process if we additionally assume that X is an aperiodic positive recurrent Harris chain under the transition kernel P.

Remark 4. If S is a discrete state space, we may simplify the conditional probabilities $R_k(x, \mathrm{d}y)$ and $R(x, \mathrm{d}y)$ somewhat. In this setting, we may view $P(x, \mathrm{d}y)$, $R_k(x, \mathrm{d}y)$, and $R(x, \mathrm{d}y)$ as matrices, and v as a vector. Then,

$$R(x, y) = \frac{v(y)P(y, x)}{\lambda v(x)},$$

$$R_k(x,y) = \frac{\boldsymbol{P}_{\tilde{\mu}}(X_{k-1}=y)P(y,x)}{\boldsymbol{P}_{\tilde{\mu}}(X_k=x)}.$$

We turn next to the development of an approximation to the distribution of X when conditioned on the event $\Gamma{=}n$. Let $\mathring{\tilde{\boldsymbol{P}}}(\cdot)$ and $\boldsymbol{P}^\circ(\cdot)$ be the probabilities on the two-sided path-space defined by

$$\mathring{\tilde{\boldsymbol{P}}}((X_j,\ldots,X_{j+m})\in\cdot) = \tilde{\boldsymbol{P}}((X_{j+1},\ldots,X_{j+m+1})\in\cdot)$$

for $j\in\mathbb{Z}$, $m\geqslant 0$, and

$$\boldsymbol{P}^\circ(X\in\cdot) = \frac{\mathring{\tilde{\boldsymbol{E}}}[I(X\in\cdot)I(X_0\in A)/h(X_{-1})]}{\mathring{\tilde{\boldsymbol{E}}}[I(X_0\in A)/h(X_{-1})]},$$

Theorem 3. *Fix* $k\geqslant 0$. *Under* A1–A3, *for* v *a.e.* x,

$$\boldsymbol{P}_x((X_{n-k},\ldots,X_{n+k})\in\cdot\,|\Gamma=n) \overset{\text{t.v.}}{\to} \boldsymbol{P}^\circ((X_{-k},\ldots,X_k)\in\cdot)$$

as $n\to\infty$.

Proof. As a consequence of (2.1), it is evident that for $n\geqslant 1$ and $x\in A$

$$\boldsymbol{P}_x(X\in\cdot\,|\Gamma=n) = \frac{\tilde{\boldsymbol{E}}_{x,n-1}[I(X\in\cdot)I(X_n\in A)/h(X_{n-1})]}{\tilde{\boldsymbol{E}}_{x,n-1}[I(X_n\in A)/h(X_{n-1})]}.$$

The rest of the argument then follows as in the proof of Theorem 1. □

We next provide some insight into the transition structure of the limit process that appears when conditioning on $\Gamma=n$ (for n large). Let

$$\mathring{w}_k(x) = \mathring{\tilde{\boldsymbol{E}}}[I(X_0\in A)/h(X_{-1})|X_k=x].$$

Theorem 4. *Assume* A1–A3. *Then*,

(i) $\boldsymbol{P}^\circ(X_{k+1}\in dy|X_j\colon j\leqslant k) = P(X_k,dy)$ *a.s. for* $k\geqslant 0$;
(ii) $\boldsymbol{P}^\circ(X_0\in dy|X_j\colon j\leqslant -1) = P(X_{-1},dy)I(y\in A)/P(X_{-1},A)$ *a.s.*
(iii) $\boldsymbol{P}^\circ(X_{k+1}\in dy|X_j\colon j\leqslant k) = G(X_k,dy)\mathring{w}_{k+1}(y)/\mathring{w}_k(X_k)$ *a.s. for* $k\leqslant -2$;
(iv) $\boldsymbol{P}^\circ(Y_{k+1}\in dy|Y_j\colon j\leqslant k) = R(Y_k,dy)$ *a.s. for* $k\geqslant 1$;
(v) $\boldsymbol{P}^\circ(Y_1\in dy|Y_j\colon j\leqslant 0) = R_1(Y_0,dy)\mathring{w}_0(y)/\mathring{w}_1(Y_0)$ *a.s.*
(vi) $\boldsymbol{P}^\circ(Y_{k+1}\in dy|Y_j\colon j\leqslant k) = R_{k+1}(Y_k,dy)\mathring{w}_k(y)/\mathring{w}_{k+1}(Y_k)$ *a.s.*

The proof is very similar to that of Theorem 2, and is therefore omitted. The two-sided limit process that appears in this setting has qualitative structure that corresponds closely to that obtained in Theorem 2. In particular, the two-sided taboo process obtained here is time-inhomogeneous before time 0 and time-homogeneous after time 0 (modulo the transition at the origin), with the time-reversed having similar behavior. Furthermore, the "forward process" X is asymptotically time-homogeneous to the left. Note that the transition probabilities of the time-reversed process after time 0 are the same as those of the time-reversed process in Theorem 2.

Remark 5. Some simplification occurs when S is discrete. In particular,

$$R_1(x,y)\frac{w_0^\circ(y)}{w_1^\circ(x)} = \frac{v(y)P(y,x)}{\sum_{z\in A} v(z)P(z,x)}.$$

Our final result in this section establishes a connection between $\boldsymbol{P}^\circ$ and $\boldsymbol{P}^*$. Let V be a geometric r.v. with mass function

$$\boldsymbol{P}(V=k)=(1-\lambda)\lambda^{k-1} \quad \text{for } k\geqslant 1 \tag{3.2}$$

and enrich the two-sided path-space so that it also supports V.

Theorem 5. *Assume* A1–A3. *Under* $\boldsymbol{P}^\circ$ *let* V *be a geometric r.v. having mass function* (3.2) *and let* V *be independent of* X. *Then, for* $k\geqslant 0$,

$$\boldsymbol{P}^*((X_{-k},\dots,X_k)\in\cdot)=\boldsymbol{P}^\circ((X_{-k-V},\dots,X_{k-V})\in\cdot).$$

Proof. For each fixed $k\geqslant 0$ and $\ell > 1$, and v a.e. x,

$$\begin{aligned}&\boldsymbol{P}_x((X_{n-k},\dots,X_{n+k})\in\cdot \quad ,\Gamma=n+\ell|\Gamma>n)\\ &\quad=\boldsymbol{P}_x((X_{n-k},\dots,X_{n+k})\in\cdot\ |\Gamma=n+\ell)\frac{\boldsymbol{P}_x(\Gamma=n+\ell)}{\boldsymbol{P}_x(\Gamma>n)}\\ &\quad\to \boldsymbol{P}^\circ((X_{-k-\ell},\dots,X_{k-\ell})\in\cdot)\boldsymbol{P}^\circ(V=\ell)\end{aligned} \tag{3.3}$$

by Proposition 1 and Theorem 3. On the other hand,

$$\begin{aligned}&\boldsymbol{P}_x((X_{n-k},\dots,X_{n+k})\in\cdot \quad ,\Gamma=n+\ell|\Gamma>n)\\ &\quad\to \boldsymbol{P}^*((X_{-k},\dots,X_k)\in\cdot \quad ,\Gamma=\ell) \quad \text{as} \quad n\to\infty.\end{aligned} \tag{3.4}$$

Equating (3.3) and (3.4), and using the independence of V and X under $\boldsymbol{P}^\circ$, we arrive at the identity

$$\begin{aligned}&\boldsymbol{P}^\circ((X_{-k-V},\dots,X_{k-V})\in\cdot \quad ,V=\ell)\\ &\quad=\boldsymbol{P}^*((X_{-k},\dots,X_k)\in\cdot \quad ,\Gamma=\ell).\end{aligned}$$

Summing over ℓ yields the theorem. □

Informally, Theorem 3.5 asserts that the two-sided limit process associated with conditioning on $\Gamma > n$ can be obtained by looking backwards a geometric amount of time in the two-sided limit process associated with conditioning on $\Gamma = n$.

4. Taboo limit theory for continuous-time Markov processes

In this section, we briefly describe the extensions of the theory of Section 3 from the discrete-time Markov chain setting to the continuous-time Markov process context. In order to streamline our arguments, we sometimes impose stronger conditions than the minimal ones required.

Let $X=(X(t)\colon t\geqslant 0)$ be a (time-homogenous) Markov process taking values in a complete separable metric space S; we assume the paths are right continuous with left-hand limits. For (measurable) $A\subseteq S$, let

$$\Gamma=\inf\{t\geqslant 0\colon X(t)\in A^{\mathrm{c}}\}$$

be the first exit time from A and assume that Γ is measurable. Let $\boldsymbol{P}_x(\cdot)$ be the probability on the path-space of X associated with conditioning on $X(0)=x\in S$. We are interested in studying the two-sided limit process that arises as an approximation to $\boldsymbol{P}((X(t+u)\colon -a\leqslant u<\infty)\in\cdot\,|\Gamma>t)$ for t large.

We start by describing the theory when S is finite and the paths have finitely many jumps in finite intervals, in which case X is a continuous-time Markov chain under $\boldsymbol{P}_x$. Let $Q=(Q(x,y)\colon x,y\in S)$ be the generator of X under $\boldsymbol{P}_x$, and let $K=(K(x,y)\colon x,y\in A)$ be the restriction of Q to A. We assume that:

A4. K is irreducible and $\boldsymbol{P}_x(\Gamma<\infty)=1$ for at least one $x\in A$.

Under A4, it follows that there exists a positive scalar c such that $c^{-1}(K+cI)$ is strictly substochastic and irreducible. The Perron–Frobenius theorem for finite matrices then implies that there exists a scalar $\beta\in(0,1)$ and positive vectors h and v such that

$$c^{-1}(K+cI)h=\beta h \quad\text{and}\quad c^{-1}v(K+cI)=\beta v,$$

from which we obtain

$$Kh=-\alpha h \quad\text{and}\quad vK=-\alpha v$$

for $\alpha=c(1-\beta)>0$. For $x,y\in A$, let

$$G(x,y)=K(x,y)\frac{h(y)}{h(x)}+\alpha\delta_{xy}$$

and observe that $G=(G(x,y)\colon x,y\in A)$ is a generator. Standard properties of matrix exponentials imply that

$$(\mathrm{e}^{Gt})(x,y)=\mathrm{e}^{\alpha t}\frac{h(y)}{h(x)}(\mathrm{e}^{Kt})(x,y).$$

Hence, if $\tilde{\boldsymbol{P}}_x(\cdot)$ is the probability on the path-space of X under which X has generator G,

$$\tilde{\boldsymbol{P}}_x(X(t)=y)=\mathrm{e}^{\alpha t}\frac{h(y)}{h(x)}\boldsymbol{P}_x(X(t)=y,\ \Gamma>t).$$

Put

$$\eta(x)=\frac{v(x)h(x)}{\sum_{z\in A}v(z)h(z)},$$

it is easily verified that $\eta=(\eta(x)\colon x\in A)$ is the stationary distribution of X under G. Let $\tilde{\boldsymbol{P}}(\cdot)$ be the probability on the two-sided path-space corresponding to $X=(X(t)\colon -\infty<t<\infty)$ under which X has finite-dimensional distributions given by

$$\begin{aligned}&\tilde{\boldsymbol{P}}(X(t_1)=x_1,\ldots,X(t_m)=x_m,X(0)=x_0,X(t_{m+1})=x_{m+1},\ldots,X(t_n)=x_n)\\ &\quad=\eta(x_1)\tilde{\boldsymbol{P}}_{x_1}(X(t_2-t_1)=x_2)\ldots\tilde{\boldsymbol{P}}_{x_m}(X(-t_m)=x_0)\\ &\qquad\boldsymbol{P}_{x_0}(X(t_{m+1})=x_{m+1})\ldots\boldsymbol{P}_{x_{n-1}}(X(t_n-t_{n-1})=x_n)\end{aligned}$$

for $t_1 < t_2 < \cdots < t_m < 0 < t_{m+1} < \cdots < t_n$ and $x_0, \dots, x_m \in A$ with $x_{m+1}, \dots x_n \in S$. Finally, let $\boldsymbol{P}^*$ be the probability on the two-sided path-space defined by

$$\boldsymbol{P}^*(X\in\cdot) = \frac{\tilde{\boldsymbol{E}}[I(X\in\cdot)/h(X(0))]}{\tilde{\boldsymbol{E}}[1/h(X(0))]}.$$

The following result has a proof essentially identical to that of Theorem 1.

Proposition 2. *Under* A4,

$$\boldsymbol{P}_x((X(t+u)\colon\ -a\leqslant u\leqslant a)\in\cdot\,|\Gamma > t) \stackrel{\text{t.v.}}{\to} \boldsymbol{P}^*((X(u)\colon\ -a\leqslant u\leqslant a)\in\cdot)$$

as $t\to\infty$, *for each* $x\in A$ *and* $a\geqslant 0$.

To describe the transition structure of X under $\boldsymbol{P}^*$, let $u(\cdot) = (u(\cdot,x)\colon x\in A)$ be the unique solution to

$$u'(t) = Gu(t),$$
$$s/t \quad u(0,x) = 1/h(x),\ x\in A.$$

Note that $u(t,x) = \tilde{\boldsymbol{E}}_x[1/h(X(t))]$. Also, let $Y = (Y(t)\colon\ -\infty < t < \infty)$ be defined by $Y(t) = X(-t)$, and set

$$v(t,x) = \sum_{z\in A} v(z)\boldsymbol{P}_z(X(t) = x)$$

for $x\in S$, $t\geqslant 0$.

Theorem 6. *Assume* A4. *Then, under* $\boldsymbol{P}^*$, X *is a Markov process for which*

$$\boldsymbol{P}^*(X(t+h) = y|X(u)\colon u\leqslant t)$$
$$= \delta_{X(t)y} + Q(t,X(t),y)h + o(h) \quad a.s.$$

as $h\downarrow 0$, *where*

$$Q(t,x,y) = G(x,y)\frac{u(-t,y)}{u(-t,x)} - \delta_{xy}\frac{u'(-t,y)}{u(-t,x)}$$

for $t<0$ *and* $x,y\in A$, *and* $Q(t,x,y) = Q(x,y)$ *for* $t\geqslant 0$ *and* $x,y\in S$. *Furthermore*, Y *is also a Markov process under* $\boldsymbol{P}^*$ *and*

$$\boldsymbol{P}^*(Y(t+h) = y|Y(u)\colon u\leqslant t)$$
$$= \delta_{Y(t)y} + Q_R(t,Y(t),y)h + o(h) \quad a.s.$$

as $h\downarrow 0$, *where*

$$Q_R(t,x,y) = Q(y,x)\frac{v(-t,y)}{v(-t,x)} - \delta_{xy}\frac{v'(-t,x)}{v(-t,x)}$$

for $t<0$ *and* $x,y\in S$, *and*

$$Q_R(t,x,y) = Q(y,x)\frac{v(y)}{v(x)} + \alpha\delta_{xy}$$

for $t\geqslant 0$ *and* $x,y\in A$.

Proof. To establish that X and Y are Markov follows the same argument as in discrete time, as does the fact that $Q(t,x,y) = Q(x,y)$ for $t \geqslant 0$. For $t < 0$, note that for $\varepsilon > 0$,

$$\begin{aligned}
&\boldsymbol{P}^*(X(t+\varepsilon) = y | X(t) = x)\\
&= \tilde{\boldsymbol{P}}_x(X(\varepsilon) = y) u(-t-\varepsilon, y)/u(-t,x)\\
&= (\delta_{xy} + G(x,y)\varepsilon + \mathrm{o}(\varepsilon))(u(-t,y) - \varepsilon u'(-t,y) + \mathrm{o}(\varepsilon)) u(-t,x)^{-1}\\
&= \delta_{xy} + \varepsilon \left\{ G(x,y) \frac{u(-t,y)}{u(-t,x)} - \delta_{xy} \frac{u'(-t,x)}{u(-t,x)} \right\} + \mathrm{o}(\varepsilon).
\end{aligned}$$

As for the reversed process, we have that for $t \geqslant 0$, $\varepsilon > 0$, and $x \neq y$,

$$\begin{aligned}
&\boldsymbol{P}^*(Y(t+\varepsilon) = y | Y(t) = x)\\
&= \tilde{\boldsymbol{P}}(X(-t-\varepsilon) = y, X(-t) = x)/\tilde{\boldsymbol{P}}(X(-t) = x)\\
&= \varepsilon \eta(y) G(y,x)/\eta(x) + \mathrm{o}(\varepsilon)\\
&= \varepsilon \frac{v(y)}{v(x)} Q(y,x) + \alpha \delta_{xy} + \mathrm{o}(\varepsilon).
\end{aligned}$$

On the other hand, for $t < 0$, $\varepsilon > 0$, and $x \neq y$, we get

$$\begin{aligned}
&\boldsymbol{P}^*(Y(t+\varepsilon) = y | Y(t) = x)\\
&= \frac{\tilde{\boldsymbol{E}}[1/h(X(0)) I(X(-t-\varepsilon) = y, X(-t) = x)]}{\tilde{\boldsymbol{E}}[1/h(X(0)) I(X(-t) = x)]}\\
&= v(-t-\varepsilon, y) \boldsymbol{P}_y(X(\varepsilon) = x)/v(-t,x)\\
&= (v(-t,y) - \varepsilon v'(t,y) + \mathrm{o}(\varepsilon))(\delta_{xy} + \varepsilon Q(y,x) + \mathrm{o}(\varepsilon)) v(-t,x)^{-1}\\
&= \delta_{xy} + \varepsilon \left\{ Q(y,x) \frac{v(-t,y)}{v(-t,x)} - \delta_{xy} \frac{v'(t,x)}{v(t,x)} \right\} + \mathrm{o}(\varepsilon)
\end{aligned}$$

as $\varepsilon \downarrow 0$, completing the proof. □

Remark 6. The limit measure $\boldsymbol{P}^*$ has exactly the same structure as in discrete time. In particular, both X and Y are time-inhomogeneous in $(-\infty, 0)$ and time-homogeneous in $(0, \infty)$. Furthermore, because of our finite state space assumption, the time-inhomogeneity to the left vanishes asymptotically.

We turn next to the two-sided limit that is appropriate to approximating the distribution of X, conditional on $\Gamma = t$ (with t large). An immediate consequence of Proposition 2 is that

$$\begin{aligned}
&\boldsymbol{P}_x((X(t+u)\colon -a \leqslant u \leqslant a) \in \cdot \,|\, \Gamma \in [t, t+h))\\
&\xrightarrow{\text{t.v.}} \boldsymbol{P}^*((X(u)\colon -a \leqslant u \leqslant a) \in \cdot \,|\, \Gamma \in [0, h))
\end{aligned}$$

as $t \to \infty$, for each $x \in A$ and $a \geqslant 0$. The next result describes the behavior of the right-hand side as we shrink h to zero. For $x, y \in S$, put $R(x,y) = -Q(x,y)/Q(x,x)$ for $x \neq y$, with $R(x,x) = 0$.

Proposition 3. *Assume* A4. *Then,*

$$\boldsymbol{P}^*(X\in\cdot\,|\Gamma\in[0,h)) \Rightarrow \boldsymbol{P}^{\circ}(X\in\cdot)$$

as $h\downarrow 0$ (*in the Skorohod topology on* $D(-\infty,\infty)$), *where*

$$\boldsymbol{P}^{\circ}(\cdot) = \boldsymbol{P}^*((X_\Gamma+t\colon\; -\infty<t<\infty)\in\cdot)$$

and has finite-dimensional distributions given by

$$\boldsymbol{P}^{\circ}(X(t_1)=x_1,\dots,X(t_m)=x_m, X(0)=x_0, X(t_{m+1})=x_{m+1},\dots,X(t_n)=x_n)$$

$$=\sum_{z\in A}\boldsymbol{P}^*(X(t_1)=x_1,\dots,X(t_m)=x_m|X(0)=z)$$

$$\times\frac{\eta(z)R(z,x_0)}{\sum_{w\in A}\sum_{y\in A^c}\eta(w)R(w,y)}$$

$$\times\boldsymbol{P}_{x_0}(X(t_{m+1})=x_{m+1},\dots,X(t_n)=x_n)$$

for $t_1<t_2<\cdots<t_m<0<t_{m+1}<\cdots<t_n$, $x_1,\dots,x_m\in A$, $x_0\in A^{\mathrm{c}}$, *and* $x_{m+1},\dots,x_n\in S$.

Proof. The weak convergence follows from Theorem 5 in Glynn and Thorisson (2000). In order to establish the finite-dimensional distributions let $\Lambda=\inf\{t\geqslant 0\colon X(t-)\neq X(0)\}$ be the first positive jump time of X and observe that

$$\boldsymbol{P}^*(X\in\cdot\,|\Gamma\in[0,h)) = \boldsymbol{P}^*(X\in\cdot\,|X(\Lambda)\in A, \Lambda\leqslant h)+\mathrm{o}(1).$$

But

$$\boldsymbol{P}^*(X(t_1)=x_1,\dots,X(t_n)=x_n|X(\Lambda)\in A,\Lambda\leqslant h)$$

$$=\sum_{\substack{z\in A\\ y\in A^{\mathrm{c}}}}\boldsymbol{P}^*(X(t_1)=x_1,\dots,X(t_n)=x_n|X(0)=x, X(\Lambda)=y,\Lambda\leqslant h)$$

$$\times\frac{\eta(z)R(z,y)}{\sum_{\substack{w\in A\\ y'\in A^{\mathrm{c}}}}\eta(w)R(w,y')}$$

$$=\sum_{z\in A}\boldsymbol{P}^*(X(t_1)=x_1,\dots,X(t_m)=x_m|X(0)=x)$$

$$\times\frac{\int_0^h\lambda(z)\exp(-\lambda(z)s)\cdot\boldsymbol{P}_{x_0}(X(t_{m+1}-s)=x_{m+1},\dots,X(t_n-s)=x_n)\,\mathrm{d}s}{1-\exp(-\lambda(z)h)}$$

$$\sum_{\substack{w\in A\\ y'\in A^{\mathrm{c}}}}\eta(w)R(w,y')$$

$$\to \sum_{z\in A} \boldsymbol{P}^*(X(t_1)=x_1,\dots,X(t_m)=x_m|X(0)=x)$$

$$\times \boldsymbol{P}_{x_0}(X(t_{m+1})=x_{m+1},\dots,X(t_n)=x_n)\frac{\eta(z)R(z,x_0)}{\sum_{\substack{w\in A\\ y\in A^c}} \eta(w)R(w,y)}$$

as $h\downarrow 0$, proving the result. □

Using the Markov structure of $\boldsymbol{P}^*$ to the left and that of $\boldsymbol{P}$ to the right, it is straightforward to describe the transition structure of X under $\boldsymbol{P}^\circ$. For reasons of brevity, we omit both the statement and proof. Furthermore, it is easy to prove a characterization result analogous to Theorem 5. Specifically, if V is an exponential (α) r.v. independent of X under $\boldsymbol{P}^\circ$, then

$$\boldsymbol{P}^*(X\in\cdot)=\boldsymbol{P}^\circ((X(u-V):\ -\infty<u<\infty)\in\cdot).$$

We conclude this section by describing the relevant theory in the diffusion context. Let $B=(B(t)\colon t\geqslant 0)$ be a k-dimensional standard Brownian motion. Suppose that A is open and that X is an $\mathbb{R}^d$-valued process that is a strong solution of the stochastic differential equation

$$\mathrm{d}X(t)=\mu(X(t))\,\mathrm{d}t+\sigma(X(t))\,\mathrm{d}B(t),$$

where $\mu:\mathbb{R}^d\to\mathbb{R}^d$ and $\sigma:\mathbb{R}^d\to\mathbb{R}^{d\times k}$ are the drift and dispersion matrices of the diffusion, respectively. For $x\in\mathbb{R}^d$, put

$$a_{i\ell}(x)=\sum_{j=1}^{k}\sigma_{ij}(x)\sigma_{\ell j}(x)$$

and let

$$L=\sum_{i=1}^{d}\mu_i(x)\frac{\partial}{\partial x_i}+\frac{1}{2}\sum_{i,j=1}^{d}a_{ij}(x)\frac{\partial^2}{\partial x_i\partial x_j}.$$

Assume that:

A5. There exists a bounded twice continuously differentiable function h such that $h(x)>0$ whenever $x\in A$ and $h(x)=0$ for $x\in A^c$ In addition, there exists a scalar $\alpha>0$ for which

$$Lh=-\alpha h.$$

Put $k(x)=\log h(x)$, and let k_n be a twice-continuously differentiable function that agrees with k on $A_n=\{x\colon h(x)>1/n\}$. Put $\Gamma_n=\inf\{t\geqslant 0\colon X(t)\in A_n^c\}$. Then, Itô's formula establishes that for $t\leqslant\Gamma_n$,

$$\mathrm{d}k_n(X(t))=(Lh)(X(t))/h(X(t))\,\mathrm{d}t+\psi(t)\,\mathrm{d}B(t)-\tfrac{1}{2}\psi(t)\psi(t)\,\mathrm{d}t \tag{4.1}$$

where $\psi(t)=(\psi_j(t)\colon 1\leqslant j\leqslant k)$ has components given by

$$\psi_j(t)=\sum_{i=1}^{d}\frac{\partial h}{\partial x_i}(X(t))\sigma_{ij}(X(t))/h(X(t)).$$

Observe, as a consequence of (4.1), that

$$\begin{aligned}&\exp(\alpha(t\wedge\Gamma_n)+k_n(X(t\wedge\Gamma_n))-k_n(X(0)))\\&\quad=\exp\left(\int_0^{t\wedge\Gamma_n}\psi(s)\,\mathrm{d}B(s)-\frac{1}{2}\int_0^{t\wedge\Gamma_n}\psi(s)\psi(s)\,\mathrm{d}s\right)\\&\quad\triangleq L(t\wedge\Gamma_n).\end{aligned}$$

Then, Girsanov's formula (see, for example, Karatsas and Shreve, 1988) asserts that if $X(0)=x\in A$, there exists a probability $\tilde{\boldsymbol{P}}_x(\cdot)$ on the path-space of X under which X satisfies, in a weak sense, the stochastic equation

$$X(t\wedge\Gamma_n)=x+\int_0^{t\wedge\Gamma_n}\tilde{\mu}(X(s))\,\mathrm{d}s+\int_0^{t\wedge\Gamma_n}\sigma(X(s))\,\mathrm{d}B(s),$$

where $\tilde{\mu}(x)=(\tilde{\mu}_i(x)\colon 1\leqslant i\leqslant d)$ is given by

$$\tilde{\mu}_i(x)=\mu_i(x)+\sum_{\ell=1}^{d}\frac{\partial h}{\partial x_\ell}(X(t))\frac{a_{i\ell}(X(t))}{h(X(t))}.$$

Furthermore,

$$\tilde{\boldsymbol{P}}_x((X(u\wedge\Gamma_n)\colon 0\leqslant u\leqslant t)\in\cdot)=\boldsymbol{E}_x[I((X(u\wedge\Gamma_n)\colon 0\leqslant u\leqslant t)\in\cdot)L(t\wedge\Gamma_n)].\tag{4.2}$$

Hence,

$$\begin{aligned}\tilde{\boldsymbol{P}}_x(\Gamma_n\leqslant t)&\leqslant\boldsymbol{E}_x[I(\Gamma_n\leqslant t)h(X(\Gamma_n))\mathrm{e}^{\alpha t}/h(x)]\\&\leqslant\frac{1}{n}\boldsymbol{P}_x(\Gamma_n\leqslant t)\mathrm{e}^{\alpha t}/h(x)\leqslant\frac{1}{n}\mathrm{e}^{\alpha t}/h(x)\to 0\end{aligned}$$

as $n\to\infty$, so that $\tilde{\boldsymbol{P}}_x(\Gamma\leqslant t)=0$ for $t\geqslant 0$. Relation (4.2) then yields

$$\boldsymbol{P}_x(\Gamma>t)=\mathrm{e}^{-\alpha t}\tilde{\boldsymbol{E}}_x[h(X(0))/h(X(t))].\tag{4.3}$$

Furthermore, (4.2) establishes that for $x\in A$,

$$\begin{aligned}&\boldsymbol{P}_x((X(u)\colon 0\leqslant u\leqslant t)\in\cdot,\Gamma>t)\\&\quad=\mathrm{e}^{-\alpha t}\tilde{\boldsymbol{E}}_x[I((X(u)\colon 0\leqslant u\leqslant t)\in\cdot)h(X(0))/h(X(t))],\end{aligned}\tag{4.4}$$

where X satisfies

$$\mathrm{d}X(t)=\tilde{\mu}(X(t))\,\mathrm{d}t+\sigma(X(t))\,\mathrm{d}B(t)\quad\text{under }\tilde{\boldsymbol{P}}_x.$$

We now add an assumption that forces X to be positive recurrent under $\tilde{\boldsymbol{P}}_x$.

A6. Suppose that there exists a probability η such that for each $x\in A$, $\tilde{\boldsymbol{P}}_x(X(t))\in\cdot)\overset{\text{t.v.}}{\to}\eta(\cdot)$ as $t\to\infty$. In addition, for $x\in A$

$$\tilde{\boldsymbol{E}}_x[1/h(X(t))]\to\int_A\eta(\mathrm{d}y)/h(y)\quad\text{as}\quad t\to\infty.$$

Let $\tilde{\boldsymbol{P}}$ be the probability on the two-sided path-space supporting $X=(X(t):-\infty<t<\infty)$ under which $\tilde{\boldsymbol{P}}(X(t)\in\cdot)=\eta(\cdot)$ for $t\leqslant 0$ and under which X satisfies

$$\mathrm{d}X(t)=b(t,X(t))\,\mathrm{d}t+\sigma(X(t))\,\mathrm{d}B(t)$$

where

$$b(t,x) = \begin{cases} \tilde{\mu}(x) & \text{for } t \leqslant 0, \\ \mu(x) & \text{for } t > 0. \end{cases}$$

Then, let $\boldsymbol{P}^*$ be the probability on the two-sided path-space under which

$$\boldsymbol{P}^*(X \in \cdot) = \frac{\tilde{\boldsymbol{E}}[I(X \in \cdot)/h(X(0))]}{\tilde{\boldsymbol{E}}[1/h(X(0))]}. \tag{4.5}$$

The following result has a proof identical to that of Theorem 1; see (4.3) and (4.4).

Proposition 4. *Assume* A5–A6. *Then,*

$$\boldsymbol{P}_x((X(t+u):\ -a \leqslant u \leqslant a) \in \cdot \,|\Gamma > t) \overset{\text{t.v.}}{\to} \boldsymbol{P}^*((X(u):\ -a \leqslant u \leqslant a) \in \cdot)$$

as $t \to \infty$, *for each* $x \in A$ *and* $a \geqslant 0$.

It follows easily from (4.5) that X is Markov under the two-sided taboo limit distribution $\boldsymbol{P}^*$. To identify the exact dynamics of X under $\boldsymbol{P}^*$, we need some additional assumptions:

A7. Put $u(t,x) = \tilde{\boldsymbol{E}}_x[1/h(X(t))]$ for $t \geqslant 0$ and $x \in A$. Then, $u = (u(t,x)$: $t \geqslant 0, x \in A)$ is a twice-continuously differentiable function satisfying $u(0,x) = 1/h(x)$ and

$$\frac{\partial u}{\partial t} = Gu,$$

where

$$G = \sum_{i=1}^{d} \tilde{\mu}_i(x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j}.$$

A8. For $1 \leqslant j \leqslant k$, let

$$\phi_j(t) = \sum_{i=1}^{d} \frac{\partial u}{\partial x_i}(-t, X(t)) \frac{\sigma_{ij}(X(t))}{u(-t, X(t))}.$$

Then, if $\phi(t) = (\phi_j(t)$: $1 \leqslant j \leqslant k)$,

$$\tilde{\boldsymbol{E}}\left[\exp\left(\frac{1}{2}\int_t^0 \phi(s) \cdot \phi(s)\,\mathrm{d}s\right)\right] < \infty \quad \text{for } t \leqslant 0.$$

Put $k(t,X(t)) = \log u(-t,X(t))$ for $t \leqslant 0$, and set $u_t(x) = u(-t,x)$. Then, Itô's formula and A7 together yield

$$\begin{aligned} \mathrm{d}k(t,X(t)) &= \left[(Gu_t)(X(t)) - \frac{\partial}{\partial t}u(-t,X(t))\right] \Big/ u(-t,X(t))\,\mathrm{d}t \\ &\quad + \phi(t)\,\mathrm{d}B(t) - \tfrac{1}{2}\phi(t) \cdot \phi(t)\,\mathrm{d}t \\ &= \phi(t)\,\mathrm{d}B(t) - \tfrac{1}{2}\phi(t) \cdot \phi(t)\,\mathrm{d}t. \end{aligned} \tag{4.6}$$

Put $b^*(t,x) = (b_i^*(t,x)$: $1 \leqslant i \leqslant d)$, where

$$b_i^*(t,x) = \begin{cases} \tilde{\mu}_i(x) + \sum_{\ell=1}^{d} \frac{\partial u}{\partial x_\ell}(-t, X(t)) \frac{a_{i\ell}(X(t))}{u(-t,X(t))}, & t \leqslant 0, \\ \mu_i(x), & t > 0. \end{cases}$$

As a consequence of (4.6) and A8, Girsanov's formula establishes that if $\boldsymbol{P}_*$ is the probability on the two-sided path-space under which

$$\mathrm{d}X(t) = b^*(t, X(t))\,\mathrm{d}t + \sigma(X(t))\,\mathrm{d}B(t),$$

then

$$\boldsymbol{P}^*((X(u)\colon -a \leqslant u \leqslant a) \in \cdot) = \tilde{\boldsymbol{E}}[I((X(u)\colon -a \leqslant u \leqslant a) \in \cdot)$$

$$\times \exp(k(0, X(0)) - k(t, X(t))] \quad \text{for } t \leqslant -a. \qquad (4.7)$$

But recall that $\exp(-k(t,X(t))) = 1/u(-t,X(t))$. Since h is bounded above, $u(-t,X(t))$ is bounded below, and consequently $\exp(-k(t,X(t))$ is bounded. Furthermore, the backwards martingale convergence theorem and A6 imply that

$$u(-t, X(t)) = \tilde{\boldsymbol{E}}[1/h(X(0)) | X(u)\colon u \leqslant t] \overset{\text{a.s.}}{\to} \tilde{\boldsymbol{E}}[1/h(X(0))]$$

as $t \to -\infty$. Hence, the Dominated Convergence Theorem can be applied to (4.7), providing the identity

$$\begin{aligned} &\boldsymbol{P}_*((X(u)\colon -a \leqslant u \leqslant a) \in \cdot) \\ &\quad = \tilde{\boldsymbol{E}}[I((X(u)\colon -a \leqslant u \leqslant a) \in \cdot) \exp(k(X(0))] / \tilde{\boldsymbol{E}}[1/h(X(0))] \\ &\quad = \frac{\tilde{\boldsymbol{E}}[I((X(u)\colon -a \leqslant u \leqslant a) \in \cdot)/h(X(0))]}{\tilde{\boldsymbol{E}}[1/h(X(0))]} \\ &\quad = \boldsymbol{P}^*((X(u)\colon -a \leqslant u \leqslant a) \in \cdot). \end{aligned}$$

In other words, $\boldsymbol{P}_* = \boldsymbol{P}^*$. We summarize this discussion with the following theorem.

Theorem 7. *Assume* A5–A8. *Then, under* $\boldsymbol{P}^*$, X *is a weak solution of*

$$\mathrm{d}X(t) = b^*(t, X(t))\,\mathrm{d}t + \sigma(X(t))\,\mathrm{d}B(t),$$

where for $1 \leqslant i \leqslant d$,

$$b_i^*(t,x) = \begin{cases} \mu_i(x) + \sum_{\ell=1}^{d} \left(\frac{(\partial h/\partial x_\ell)(X(t))}{h(X(t))} + \frac{(\partial u/\partial x_\ell)(-t, X(t))}{u(-t, X(t))} \right) a_{i\ell}(X(t)), & t \leqslant 0 \\ \mu_i(x), & t > 0. \end{cases}$$

Remark 7. As in the continuous-time Markov chain setting, it is straightforward to show that $b_i^*(\cdot, x)$ is continuous at $t = 0$ for $x \in A$.

For the development of a result in the diffusion setting comparable to Proposition 3, we refer the reader to the general theory provided in our companion paper; Glynn and Thorisson (2000).

5. Coupling from the past

In this section, we provide an extension of the idea of Propp and Wilson (1996) to our current context. Specifically, we are concerned here with exact sampling from a taboo-stationary version of a discrete- or continuous-time Markov chain for which A is a finite set of states. Somewhat remarkably, the algorithm presented below provides exact samples of the taboo-stationary version without assuming any apriori knowledge of the solution to the eigenvalue problem defined in A1. In particular, the algorithm assumes no knowledge of the Perron–Frobenius eigenvalue λ, or its corresponding Perron–Frobenius eigenfunction h and eigenmeasure v.

Let $X=(X_t\colon t\geqslant 0)$ be a discrete- or continuous-time Markov chain and let Γ be the first exit time out of a finite irreducible aperiodic set of states A. Let $X^*=(X_t^*\colon -\infty<t<\infty)$ be the two-sided taboo limit process, that is, in the discrete time case X^* has the distribution $\boldsymbol{P}^*$ in Theorem 1 and in the continuous time case X^* has the distribution $\boldsymbol{P}^*$ in Proposition 2. Fix an $m\geqslant 1$.

Initial step: Fix $x\in A$ and generate i.i.d. versions $X^{(-m,x,1)}$, $X^{(-m,x,2)}$, $X^{(-m,x,3)},\ldots$ of X starting at time $-m$ in state x. Let $\Gamma^{(-m,x,1)}$, $\Gamma^{(-m,x,2)}$, $\Gamma^{(-m,x,3)},\ldots$ be the first exit times out of A and continue generating until

$$K^{(-m,x)}=\inf\{k\geqslant 1\colon \Gamma^{(-m,x,k)}>0\}.$$

Put

$$X^{(-m,x)}{:=}X^{(-m,x,K^{(-m,x)})}$$

and note that

$$\boldsymbol{P}(X^{(-m,x)}\in\cdot)=\boldsymbol{P}_x((X_{m+t}\colon -m\leqslant t<\infty)\in\cdot\,|\Gamma>m).$$

Do this for each $x\in A$.

Recursive steps: For $n>m$, fix $x\in A$ and generate i.i.d. versions $X^{(-n,x,1)}$, $X^{(-n,x,2)}$, $X^{(-n,x,3)},\ldots$ of X starting at time $-n$ in state x. Let $\Gamma^{(-n,x,1)}$, $\Gamma^{(-n,x,2)}$, $\Gamma^{(-n,x,3)},\ldots$ be the first exit times out of A and continue generating until

$$K^{(-n,x)}=\inf\{k\geqslant 1\colon \Gamma^{(-n,x,k)}>0\}.$$

Define a chain $X^{(-n,x)}$ by

$$X^{(-n,x)}=(x,X^{(-(n-1),y)})\quad \text{if } X_{-(n-1)}^{(-n,x,K^{(-n,x)})}=y,\quad y\in A,$$

and note that

$$\boldsymbol{P}(X^{(-n,x)}\in\cdot)=\boldsymbol{P}_x((X_{n+t}\colon -n\leqslant t<\infty)\in\cdot\,|\Gamma>n). \tag{5.1}$$

Do this for each $x\in A$.

Final step: Continue until

$$N=\inf\{n\geqslant m\colon \text{ the set } \{X_{-m}^{(-n,x)}\colon x\in A\} \text{ is a singleton}\}.$$

Theorem 8. *Let X be a discrete-time or continuous-time Markov chain and A a finite irreducible aperiodic set of states. Let $m\geqslant 0$ and $x\in A$. Then N is finite a.s. and*

$$(X_t^{(-N,x)}\colon -m\leqslant t<\infty)\quad \text{is a copy of}\quad (X_t^*\colon -m\leqslant t<\infty). \tag{5.2}$$

Proof. According to Remark 8 below, there are x_0, n_0, n_1 and $p > 0$ such that

$$\boldsymbol{P}_x(X_{n_0} = x_0 | \Gamma > n) \geqslant p \quad \text{for all } x \in A \text{ and } n \geqslant n_1. \tag{5.3}$$

We may take $n_1 \geqslant m$. The events

$$B_k = \{\text{the random set } \{X_{-n_1 - kn_0 + n_0}^{(-n_1 - kn_0, x)}: x \in A\} \text{ is a singleton}\}, \ k \geqslant 1,$$

are independent and

$$\boldsymbol{P}(B_k) \geqslant \boldsymbol{P}(X_{-n_1 - kn_0 + n_0}^{(-n_1 - kn_0, x, K^{(-n_1 - kn_0, x)})} = x_0 \text{ for all } x \in A)$$
$$\geqslant p^{\# A} =: q > 0.$$

Thus,

$$\boldsymbol{P}(N > n_1 + kn_0) \leqslant (1 - q)^k$$

and sending k to ∞ shows that $\boldsymbol{P}(N < \infty) = 1$.

It remains to establish (5.2). We have

$$(X_t^{(-n,x)}: \ -m \leqslant t < \infty) = (X_t^{(-N,x)}: \ -m \leqslant t < \infty) \quad \text{on } \{N \leqslant n\}$$

and thus, as $n \to \infty$,

$$\boldsymbol{P}((X_t^{(-n,x)}: \ -m \leqslant t < \infty) \in \cdot) \xrightarrow{\text{t.v.}} \boldsymbol{P}((X_t^{(-N,x)}: \ -m \leqslant t < \infty) \in \cdot).$$

From this and (5.1) it follows that (5.2) holds. □

Remark 8. In the above proof we needed the result that there are x_0, n_0, n_1 and $p > 0$ such that (5.3) holds. For a discrete-time X this is a straightforward consequence of the next proposition. For a continuous-time X this in turn follows from the fact that X observed at integer times forms a discrete-time chain and that A is irreducible aperiodic for this discrete-time chain if it is irreducible for X.

Proposition 5. *Consider a discrete-time Markov chain and suppose S is discrete and* A1–A3 *are in force. Then, for $m \geqslant 0$, $x, y \in A$,*

$$\boldsymbol{P}_x(X_m = y | \Gamma > n + m) \to G^m(x, y) \quad \text{as } n \to \infty,$$

where $G^m = (G^m(x, y): x, y \in A)$ is the mth power of G.

Proof. Note that (3.1) implies that

$$\boldsymbol{P}(X_m = y | \Gamma > n + m) = \frac{\tilde{\boldsymbol{E}}_{x,n+m}[I(X_m = y)/h(X_{n+m})]}{\tilde{\boldsymbol{E}}_{x,n+m}[1/h(X_{n+m})]}$$
$$= \frac{G^m(x, y) \sum_{z \in A} G^n(y, z)/h(z)}{\sum_{z \in A} G^{n+m}(x, z)/h(z)} \to G^m(x, y)$$

as $n \to \infty$. □

References

Brémaud, P., 1980. Point Processes and Queues: Martingale Dynamics. Springer, New York.

Chung, K.L., 1974. A Course in Probability Theory. Academic Press, New York.

Ferrari, P., Kesten, H., Martinez, S., 1996. *R*-positivity, quasi-stationary distributions and ratio limit theorems for a class of probabilistic automata. Ann. Appl. Probab. 6, 577–616.

Glynn, P.W., Thorisson, H., 2000. Taboo-stationarity and limit theory for taboo-regenerative processes. Submitted for publication.

Karatsas, I., Shreve, S., 1988. Brownian Motion and Stochastic Calculus. Springer, New York.

Karlin, S., Taylor, H., 1975. A First Course in Stochastic Processes. Academic Press, New York.

Meyn, S., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Springer, New York.

Ney, P., Nummelin, E., 1987. Markov additive processes I. Eigenvalue properties and limit theorems. Ann. Probab. 15, 561–592.

Nummelin, E., 1984. General Irreducible Markov Chains and Non-Negative Operators. Cambridge University Press, Cambridge.

Nummelin, E., Arjas, E., 1976. A direct construction of the *R*-invariant measure for a Markov chain on a general state space. Ann. Probab. 4, 674–679.

Nummelin, E., Tweedie, R., 1978. Geometric ergodicity and *R*-positivity for general Markov chains. Ann. Probab. 6, 404–420.

Propp, J.G., Wilson, D.B., 1996. Exact sampling with coupled Markov chains and applications to statistical mechanics. Random Struct. Algorithms 9, 223–252.

Seneta, E., Vere-Jones, D., 1966. On quasi-stationary distributions in discrete-time Markov chains with a denumerable infinity of states. J. Appl. Probab. 3, 403–434.

Thorisson, H., 2000. Coupling, Stationarity, and Regeneration. Springer, New York.

Tweedie, R., 1974. Quasi-stationary distributions for Markov chains on a general state space. J. Appl. Probab. 11, 726–741.