

IMPORTANCE SAMPLING TECHNIQUES FOR THE MULTIDIMENSIONAL RUIN PROBLEM FOR GENERAL MARKOV ADDITIVE SEQUENCES OF RANDOM VECTORS

BY J. F. COLLAMORE

ETH Zürich

Let $\{(X_n, S_n) : n = 0, 1, \dots\}$ be a Markov additive process, where $\{X_n\}$ is a Markov chain on a general state space and S_n is an additive component on \mathbb{R}^d . We consider $\mathbf{P}\{S_n \in A/\varepsilon, \text{ some } n\}$ as $\varepsilon \rightarrow 0$, where $A \subset \mathbb{R}^d$ is open and the mean drift of $\{S_n\}$ is away from A . Our main objective is to study the *simulation* of $\mathbf{P}\{S_n \in A/\varepsilon, \text{ some } n\}$ using the Monte Carlo technique of importance sampling. If the set A is convex, then we establish (i) the precise dependence (as $\varepsilon \rightarrow 0$) of the estimator variance on the choice of the simulation distribution and (ii) the existence of a *unique* simulation distribution which is efficient and optimal in the asymptotic sense of D. Siegmund [*Ann. Statist.* **4** (1976) 673–684]. We then extend our techniques to the case where A is not convex. Our results lead to positive conclusions which complement the multidimensional counterexamples of P. Glasserman and Y. Wang [*Ann. Appl. Probab.* **7** (1997) 731–746].

1. Introduction. There has been much recent interest in developing simulation techniques for estimating “rare event” probabilities; formally, these are probabilities $\mathbf{P}(C_\varepsilon)$, for small ε , where $\mathbf{P}(C_\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$. When direct Monte Carlo methods are used to estimate such small probabilities, one runs a numerical experiment involving n trials and computes the proportion of times that the event C_ε occurs. However, under this direct approach, one then obtains that the *error* of the estimate $\nearrow \infty$ as $\varepsilon \rightarrow 0$ when compared with the estimate [cf. Asmussen (1999), page 45]. The subject of rare event simulation consequently deals with alternative methods for simulating such probabilities which, in various contexts, are efficient and which *remain* effective in the asymptotic limit as $\varepsilon \rightarrow 0$. In many one-dimensional problems, mainly involving i.i.d. sums which, in the event C_ε , attain some region in \mathbb{R}^1 , effective simulation techniques are well known; however, in higher dimensions—as we shall soon explain—the situation is somewhat more complicated.

The objective of this article will be to study rare event simulation in the context of the following *multidimensional* boundary crossing problem: Let S_1, S_2, \dots be a sequence of random variables in \mathbb{R}^d , and consider the hitting probability of a

Received August 2000; revised July 2001.

AMS 2000 subject classifications. Primary 65C05; secondary 65U05, 60F10, 60J15, 60K10.

Key words and phrases. Monte Carlo methods, rare event simulation, hitting probabilities, large deviations, Harris recurrent Markov chains, convex analysis.

region $A/\varepsilon = \{x/\varepsilon : x \in A\}$ by $\{S_n\}$; namely, consider

$$(1.1) \quad \mathbf{P}\left\{S_n \in \frac{A}{\varepsilon}, \text{ some } n\right\} = \mathbf{P}\{T^\varepsilon(A) < \infty\} \quad \text{as } \varepsilon \rightarrow 0,$$

where

$$(1.2) \quad T^\varepsilon(A) = \inf\left\{n : S_n \in \frac{A}{\varepsilon}\right\}.$$

It will be assumed that the mean drift of $\{S_n\}$ is directed away from A , so that the probabilities in (1.1) will tend to zero as $\varepsilon \rightarrow 0$. Our objective will be to develop a numerical regime based on *importance sampling* which yields an efficient estimate for (1.1), for any fixed ε , and which has certain optimality properties as $\varepsilon \rightarrow 0$.

Analytically, the first work on problems of this type seems to have appeared in Lundberg (1909). Here, a stochastic model for the capital fluctuations of an insurance company was introduced, and the risk faced by a company under this model was studied. Under Lundberg's model, an insurance company gains capital from a constant stream of premiums inflow, and loses capital as a result of i.i.d. claims arising at a Poisson rate. These assumptions imply that the total capital gain by time t , denoted S_t , is a Lévy process. Assuming that this process has positive drift, the *ruin problem* then considers $\mathbf{P}\{S_t < -1/\varepsilon, \text{ some } t \geq 0\}$, that is, the probability that a company with an initial capital of $1/\varepsilon$ will *ever* have negative total capital, or incur ruin. A classical result due to Cramér (1930) states

$$(1.3) \quad \mathbf{P}\left\{S_t < -\frac{1}{\varepsilon}, \text{ some } t \geq 0\right\} \sim C e^{-R/\varepsilon} \quad \text{as } \varepsilon \rightarrow 0$$

for certain constants C and R .

Cramér's result and techniques were later extended to more general processes and adapted in other applied areas, such as queueing theory and sequential analysis. An extension of (1.3) to higher dimensions was given in Collamore (1996a, b). There it was shown that if A is an arbitrary open subset of \mathbb{R}^d and S_1, S_2, \dots are the sums of an i.i.d., Markov or more general sequence of random variables, then

$$(1.4) \quad \lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{P}\{T^\varepsilon(A) < \infty\} = - \inf_{v \in A} I_{\mathcal{P}}(v),$$

where $I_{\mathcal{P}}$ is the support function of the d -dimensional surface $\{\alpha : \Lambda_{\mathcal{P}}(\alpha) \leq 0\}$ and $\Lambda_{\mathcal{P}}$ is the cumulant generating function of $\{S_n/n\}$. Further distributional properties of $T^\varepsilon(A)$ were explored in Collamore (1998). This multidimensional problem is relevant, for example, to applications in insurance and finance, where it is often of interest to measure risk which is associated with several *dependent* capital factors. Also, (1.4) serves as a preliminary study for certain queueing network problems that can be modelled as a reflected random walk in \mathbb{R}^d [cf. Borovkov and Mogulskii (1996)].

While (1.3) and (1.4) provide useful asymptotic results, which [for (1.3)] may be quite accurate when the limit is removed from the left-hand side, these estimates give no indication about the rate at which the convergence to the limit actually takes place. To circumvent this problem in a closely related problem in sequential analysis, Siegmund (1976) introduced the numerical technique of importance sampling [cf. Hammersley and Handscomb (1964), pages 57–59]. According to this technique, one observes by the Radon–Nikodym theorem that

$$(1.5) \quad \mathbf{P}(C) = \mathbf{E}_{\mathbf{Q}} \left(\frac{d\mathbf{P}}{d\mathbf{Q}}(Z) \mathbb{1}_C(Z) \right),$$

where $Z \sim \mathbf{Q}$, $\mathbb{1}_C =$ indicator function on C .

Then $\mathbf{P}(C)$ is computed by simulating $\mathcal{E} \stackrel{\text{def}}{=} \frac{d\mathbf{P}}{d\mathbf{Q}}(Z) \mathbb{1}_C(Z)$ under the distribution \mathbf{Q} and averaging the empirical samples of \mathcal{E} . In the context of the standard two-sided boundary crossing problem in sequential analysis, Siegmund showed that a judicious choice of \mathbf{Q} leads to a much-reduced variance for the estimator \mathcal{E} compared with direct Monte Carlo simulation. Moreover, he showed that there is a *unique* choice of \mathbf{Q} which, in an appropriate asymptotic sense, is optimal. Extensions of Siegmund’s algorithm to other large deviations problems in \mathbb{R}^1 were later given, for example, in Asmussen (1989), Lehtonen and Nyrhinen (1992a, b) and Bucklew, Ney and Sadowsky (1990). Related developments can also be found, for example, in Cottrell, Fort and Malgouyres (1983), Chen, Lu, Sadowsky and Yao (1993), Sadowsky (1993) and Bucklew (1998).

The difficulty of extending Siegmund’s algorithm beyond the one-dimensional setting was documented by Glasserman and Wang (1997). Here it was shown that there is *no hope* of obtaining results like Siegmund’s for the multidimensional problem in (1.1) when $A = \{(u, v) : u > 1 \text{ or } v > 1\} \subset \mathbb{R}^2$ and $\mathbf{E}(S_1) = -(\mu_1, \mu_2)$, where $\mu_i > 0$. In particular, from a reasonable class of importance sampling regimes, they showed that various regimes lead to unbounded relative error for the estimator as $\varepsilon \rightarrow 0$. Similar “counterexamples” in a queueing context were given in Glasserman and Kou (1995). These counterexamples show that the much-used technique of minimizing the variational formula in Mogulskii’s theorem [Dembo and Zeitouni (1998), Theorem 5.1] does *not* lead to any sort of efficient simulation regime in general. In this paper, we will show that if the set $A \subset \mathbb{R}^d$ in (1.1) is *convex*, then a natural analog of Siegmund’s algorithm can be developed for the multidimensional problem in (1.1).

A related work is Sadowsky (1996), which established certain necessary and sufficient conditions for efficient importance sampling of $\mathbf{P}\{S_n/n \in A\}$, where A is a subset of a Banach space. A sufficient condition is established under hypotheses which roughly amount to the existence of a “dominating point” [see Ney (1983), where dominating points are defined and their existence is studied]. There are, however, some substantial differences between the results and general approach of Sadowsky’s paper and those given here. First, our problem and techniques

are very different; for example, the rate function $I_{\mathcal{P}}$ in (1.4) is neither strictly convex nor differentiable, nor are its level sets compact. Hence our analysis is quite distinct from that used for sample means or other classical large deviations theorems, such as, for example, Mogulskii’s theorem, and the rate functions and optimal importance sampling regimes we obtain are, of course, different. Second, Sadowsky works with arbitrary sequences satisfying the large deviation principle. In practice, one would usually like to simulate using an explicit transformation on the *increments* $S_1 - S_0, S_2 - S_1, \dots$, rather than on S_n itself, and it is not evident when this can be done in the framework of general sequences. Thus, an additional objective of this paper is to propose a wide and very natural class of processes—the Markov additive processes in general state space—where such explicit transformations can be achieved. (In our problem, it appears that such transformations *cannot* be obtained for the more general “Gärtner–Ellis” sequences.)

We now turn to a more precise statement of our results. Let $\{(X_n, \xi_n) : n = 0, 1, \dots\}$ be a Markov chain on $\mathbb{S} \times \mathbb{R}^d$, where \mathbb{S} is a general state space, with transition kernel of the form

$$\mathbf{P}\{(X_{n+1}, \xi_{n+1}) \in E \times \Gamma \mid X_n = x\} = \mathcal{P}(x, E \times \Gamma).$$

We are interested in the behavior of the sums $S_n = \xi_1 + \dots + \xi_n, n = 1, 2, \dots$ ($S_0 = 0$). The process $\{(X_n, S_n) : n = 0, 1, \dots\}$ is a *Markov additive process*. The simplest examples are when S_1, S_2, \dots are the sums of an i.i.d. sequence of random variables, or the sums of functions of a finite state Markov chain. More generally, we may take $\xi_{n+1} = F(X_n, X_{n+1})$, where $\{X_n\}$ is a Markov chain on a general state space, \mathbb{S} , and $F(x, y)$ is a deterministic or random function for any given $x, y \in \mathbb{S}$. Our objective is to estimate

$$\mathbf{P}\{T^\varepsilon(A) < \infty\} = \mathbf{P}\left\{S_n \in \frac{A}{\varepsilon}, \text{ some } n \in \mathbb{Z}_+\right\}$$

using importance sampling.

The importance sampling technique suggests that we simulate $\mathbf{P}\{T^\varepsilon(A) < \infty\}$ using *another* Markov additive sequence $\{(\tilde{X}_n, \tilde{S}_n) : n = 0, 1, \dots\}$ having transition kernel $\mathcal{Q} = \mathbf{P}\{(\tilde{X}_{n+1}, \tilde{\xi}_{n+1}) \in E \times \Gamma \mid \tilde{X}_n = x\}$. An adaptation of (1.5) then becomes

$$\mathbf{P}\{T^\varepsilon(A) < \infty\} = \mathbf{E}_{\mathcal{Q}}(\mathcal{E}_{\mathcal{Q}, \varepsilon})$$

for some “estimator” $\mathcal{E}_{\mathcal{Q}, \varepsilon}$ that is computed from the \mathcal{Q} -distributed sequence of simulated random variables $\{X_0, \dots, X_{T^\varepsilon(A)}; S_0, \dots, S_{T^\varepsilon(A)}\}$. The main objective is to choose \mathcal{Q} so that it minimizes $\text{Var}_{\mathcal{Q}}(\mathcal{E}_{\mathcal{Q}, \varepsilon})$ as $\varepsilon \rightarrow 0$, or equivalently $\mathbf{E}_{\mathcal{Q}}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2)$ as $\varepsilon \rightarrow 0$.

Under the assumption that A is convex, our first result provides a large deviations estimate of the form

$$(1.6) \quad \lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}_{\mathcal{Q}}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2) = - \inf_{v \in \mathfrak{A}} I_{\mathcal{K}_{\mathcal{Q}}}(v),$$

for some subset \mathfrak{A} of ∂A and some “rate function” $I_{\mathcal{K}_{\mathcal{Q}}}$. This establishes the *precise* correspondence between \mathcal{Q} and the decay (or growth) rate of $\mathbf{E}(\mathcal{E}_{\mathcal{Q},\varepsilon}^2)$ as $\varepsilon \rightarrow 0$. The implication of this estimate is made clear in Example 3.1, where the level sets of $f(\mathcal{Q}) \stackrel{\text{def}}{=} \mathbf{E}(\mathcal{E}_{\mathcal{Q},\varepsilon}^2)$ are described by an explicit asymptotic formula. [For sample means of a uniformly recurrent Markov chain in \mathbb{R} , a related “second-moment” estimate was established in Bucklew, Ney and Sadowsky (1990).]

From (1.6) we may draw several conclusions. Equation (1.6) provides, for example, information about the robustness of a simulation regime under small perturbations of \mathcal{Q} . More important, (1.6) shows that an efficient simulation regime should be one which maximizes (over all simulation distributions \mathcal{Q}) the rate function appearing on the right-hand side, namely, $J(\mathcal{Q}) \stackrel{\text{def}}{=} \inf_{v \in \mathfrak{A}} I_{\mathcal{K}_{\mathcal{Q}}}(v)$. If A is convex, then we show that there exists a *unique* choice of \mathcal{Q} which maximizes $J(\mathcal{Q})$. Furthermore, under this optimal distribution we show that simulation is indeed efficient in the sense that it has “logarithmic efficiency” and very often “bounded relative error;” this will be achieved using a convexity result given below in Lemma 3.2. It will be shown that the optimality we obtain is quite general and extends to the case where the simulation distribution is allowed to be time-dependent.

Finally, we show that if A is a general set, then it is possible to partition A into a finite subcollection, A_1, \dots, A_l , and simulate independently along the elements of this subpartition. We show that a useful partition can *always* be obtained. The basic idea is to partition A along the *level sets* of the function $I_{\mathcal{P}}$ in (1.4). The estimator we obtain will generally be efficient and in some main cases will have “bounded relative error.” [All of the above results can easily be generalized to finite time-horizon problems of the form $\mathbf{P}\{T^\varepsilon(A) < K/\varepsilon\}$, $K < \infty$; the required modifications follow along the lines of Collamore (1998).]

We will establish our results in some generality, at the level of Markov additive processes in general state space, as studied in a large deviations context by Ney and Nummelin (1987a, b), de Acosta (1988), de Acosta and Ney (1998) and references therein, following along the lines of the seminal papers of Donsker and Varadhan (1975, 1976, 1983). Thus, our results differ from known importance sampling results given, for example, in Siegmund (1976), Asmussen (1989) and Lehtonen and Nyrhinen (1992a, b), which focus on i.i.d. sums or the sums of a finite state space Markov chain, and in Bucklew, Ney and Sadowsky (1990), where sums of a general state space Markov chain are considered, but under a strong uniform recurrence condition. The usefulness of this general approach is illustrated in Example 3.2, where our results are applied to the ARMA(p, q) time series models. These models can be viewed as Markov additive processes in general state space, but they do not satisfy the uniform recurrence condition assumed, for example, in Bucklew, Ney and Sadowsky (1990). The mathematical difficulty inherent in the study of general Markov additive processes lies in the absence of a corresponding Perron–Frobenius theory, which is the basis for the analysis of finite or uniformly

recurrent Markov chains. Our extension to general Markov additive processes will be achieved using the theory of nonnegative operators, as given in Nummelin (1984); therefore our approach is similar to that of Ney and Nummelin (1987a, b). However, in our case, we will make use of abstract renewal properties, and—in contrast with Ney and Nummelin’s work—our renewal structure will *not* generally coincide with the inherent renewal structure of the Markov additive process or of the simulated process.

In the next section, we introduce Markov additive processes in general state space and provide some necessary background on these processes and on nonnegative kernels. The main results are stated formally in Section 3 and proved in Section 4.

2. Background.

2.1. Markov additive processes: definition and regenerative property. Let $\{X_n : n = 0, 1, \dots\}$ be a Markov chain on a countably generated general measurable space $(\mathbb{S}, \mathcal{S})$. Assume $\{X_n\}$ is aperiodic and irreducible with respect to a maximal irreducibility measure φ .

To this Markov chain adjoin a sequence $\{\xi_n\}$ such that $\{(X_n, \xi_n) : n = 0, 1, \dots\}$ is a Markov chain on $(\mathbb{S} \times \mathbb{R}^d, \mathcal{S} \times \mathcal{R}^d)$ with transition kernel

$$(2.1) \quad \mathcal{P}(x, E \times \Gamma) \stackrel{\text{def}}{=} \mathbf{P}\{(X_{n+1}, \xi_{n+1}) \in E \times \Gamma \mid X_n = x\},$$

for all $x \in \mathbb{S}$, $E \in \mathcal{S}$, $\Gamma \in \mathcal{R}^d$, where \mathcal{R}^d denotes the Borel σ -algebra on \mathbb{R}^d . Let \mathfrak{F}_n denote the σ -algebra generated by $\{X_0, \dots, X_n, S_0, \dots, S_n\}$; and let $S_n = \xi_1 + \dots + \xi_n$, $n = 1, 2, \dots$, and $S_0 = 0$. The sequence $\{(X_n, S_n) : n = 0, 1, \dots\}$ is a Markov additive process.

A φ -irreducible Markov chain *always* has a minorization [Nummelin (1984), Theorem 2.1]. Following Ney and Nummelin (1987a, b), we will work with a hypothesis which extends this minorization to Markov additive processes.

Minorization.

(M) For some family of measures $\{h(x, \Gamma) : \Gamma \in \mathcal{R}^d\}$ on \mathbb{R}^d , $x \in \mathbb{S}$, and some probability measure $\{v(E \times \Gamma) : E \in \mathcal{S}, \Gamma \in \mathcal{R}^d\}$ on $\mathbb{S} \times \mathbb{R}^d$,

$$h(x, \cdot) * v(E \times \cdot) \leq \mathcal{P}(x, E \times \Gamma) \quad \text{for all } x \in \mathbb{S}, E \in \mathcal{S}, \Gamma \in \mathcal{R}^d.$$

(The symbol $*$ denotes convolution. We will often abbreviate the left-hand side by $h \otimes v$.) As in Ney and Nummelin (1987a, b), we will generally assume that either $v(dy \times ds) = v(dy)\eta_\emptyset(ds)$ or $h(x, ds) = h(x)\eta_\emptyset(ds)$, where η_\emptyset denotes a measure on \mathbb{R}^d having point mass at the origin; in other words, we will assume the slightly stronger condition:

(\mathfrak{M}) One of the following minorizations holds:

$$h(x, \Gamma)v(E) \leq \mathbf{P}(x, E \times \Gamma) \quad \text{or} \quad h(x)v(E \times \Gamma) \leq \mathbf{P}(x, E \times \Gamma),$$

where $\{h(\cdot)\}$ is a family of measures on \mathbb{R}^d , for each $x \in \mathbb{S}$ (respectively, a function on \mathbb{S}), and $v(\cdot)$ is a probability measure on \mathbb{S} (respectively, on $\mathbb{S} \times \mathbb{R}^d$).

In a few situations, we will strengthen this minorization to the following.

(\mathfrak{R}) $av(E \times \Gamma) \leq \mathcal{P}(x, E \times \Gamma) \leq bv(E \times \Gamma)$, for all $x \in \mathcal{S}$, $E \in \mathcal{S}$, $\Gamma \in \mathcal{R}^d$, where v is as in (\mathfrak{M}), and a, b are positive constants.

When (\mathfrak{R}) holds, the Markov additive process is said to be “uniformly recurrent.” This strong recurrence condition is satisfied, for example, if S_n denotes the sums of functions of a finite state Markov chain, but it is often not satisfied for more general processes, such as the AR(p) processes of Example 3.2 below. We emphasize, however, that condition (\mathfrak{R}) will *not* be required for the main results of this paper given below in Theorems 3.1 and 3.3.

Under (\mathfrak{M}), a regenerative structure can be deduced for the Markov additive process:

LEMMA 2.1. *Let $\{(X_n, S_n)\}_{n \geq 0}$ be a Markov additive process satisfying (\mathfrak{M}). Then there exist random variables $0 < T_0 < T_1 < \dots$ and a decomposition $\xi_{T_i} = \xi'_{T_i} + \xi''_{T_i}$, $i = 0, 1, \dots$, with the following properties:*

- (i) $\{T_{i+1} - T_i : i = 0, 1, \dots\}$ are i.i.d. and finite a.e.;
- (ii) the random blocks $\{X_{T_i}, \dots, X_{T_{i+1}-1}, \xi''_{T_i}, \xi_{T_{i+1}}, \dots, \xi_{T_{i+1}-1}, \xi'_{T_{i+1}}\}$ are independent;
- (iii) $\mathbf{P}_x\{(X_{T_i}, \xi''_{T_i}) \in E \times \Gamma'' \mid \mathfrak{F}_{T_i-1}, \xi'_{T_i}\} = v(A \times \Gamma'')$ for all $E \in \mathcal{S}$ and $\Gamma'' \in \mathcal{R}^d$.

For Harris recurrent Markov chains, this lemma was established by Athreya and Ney (1978) and Nummelin (1978). The extension to Markov additive processes is in Ney and Nummelin (1984).

REMARK 2.1. (i) If the function h in (\mathfrak{M}) is independent of x , that is, if the lower bound of (\mathfrak{R}) holds, then $\mathbf{P}\{T_i = n, \text{ some } i \mid \mathfrak{F}_{n-1}\} \geq a$, where a is the positive constant in (\mathfrak{R}). Thus, in particular, $\mathbf{E}(T_{i+1} - T_i) < \infty$, $i = 0, 1, \dots$, and $\mathbf{E}(T_0) < \infty$.

(ii) If $h(x, ds) = h(x)\eta_\emptyset(ds)$, then $\xi'_{T_i} = 0$, $i = 0, 1, \dots$. If $v(dy \times ds) = v(dy)\eta_\emptyset(ds)$, then $\xi''_{T_i} = 0$, $i = 0, 1, \dots$. [See Ney and Nummelin (1984).]

Futher properties of Markov chains in general state space can be found in Nummelin (1984), Revuz (1975) and Meyn and Tweedie (1993). Further properties of Markov additive processes can be found in the large deviations papers of de Acosta (1988), de Acosta and Ney (1998) and especially Ney and Nummelin (1987a, b).

2.2. *Nonnegative kernels, eigenvalues and eigenvectors.* We will also need certain facts about nonnegative kernels, which we now summarize and apply in the context of Markov additive processes. For details, see Nummelin (1984).

Let $\{K(x, E) : x \in \mathbb{S}, E \in \mathcal{E}\}$ be a σ -finite nonnegative φ -irreducible kernel on a countably generated measurable space $(\mathbb{S}, \mathcal{E})$. For any function $h : \mathbb{S} \rightarrow \mathbb{R}$ and any measure ν on $(\mathbb{S}, \mathcal{E})$, let

$$Kh(x) = \int K(x, dy)h(y), \quad \nu K(E) = \int \nu(dx)K(x, E),$$

$$(h \otimes \nu)(x, E) = h(x)\nu(E), \quad \nu h(E) = \int_E \nu(dx)h(x), \quad \nu h = \nu h(\mathbb{S}).$$

Assume

$$(2.2) \quad h \otimes \nu \leq K.$$

Define

$$G^{(\rho)} = \sum_{n=0}^{\infty} \rho^n K^n, \quad G_{h,\nu}^{(\rho)} = \sum_{n=0}^{\infty} \rho^n (K - h \otimes \nu)^n,$$

$$b_n = \nu(K - h \otimes \nu)^{n-1}h, \quad \hat{b}(\rho) = \sum_{n=1}^{\infty} \rho^n b_n.$$

For any irreducible kernel K , there exists a constant R such that $0 \leq R < \infty$, and $G^{(\rho)}$ is “finite” for $\rho < R$ and “infinite” for $\rho > R$ [Nummelin (1984), pages 27–28]. The constant R is called the *convergence parameter* of K . A kernel K with convergence parameter R is said to be *R-recurrent* if $G^{(R)}(x, E) = \infty$ for $x \in \mathbb{S}$, $\varphi(E) > 0$, and *R-transient* if this is not true. It can be shown that K is *R-recurrent* if and only if $\hat{b}(R) = 1$.

A function $r : \mathbb{S} \rightarrow [0, \infty]$ (not $\equiv \infty$) is *ρ -subinvariant* if $\rho Kr \leq r$, and *invariant* (with unique eigenvalue $\lambda = \rho^{-1}$) if $\rho Kr = r$. If $R > 0$ is the convergence parameter of K , then the existence of invariant and subinvariant functions for K can be obtained under (2.2), as follows. If $\rho < R$ or if $\rho = R$ and K is *R-transient*, then a ρ -subinvariant function exists [given by $r(x) = (G^{(\rho)}h)(x)$]. If K is *R-recurrent*, then an *R*-invariant function exists [given by $r(x) = (RG_{h,\nu}^{(R)}h)(x)$]. [See Nummelin (1984), Proposition 5.2 and Theorem 5.1.]

Now specialize to the transformed Markov additive kernel $\hat{\mathcal{P}}(\alpha)$, where (for any kernel K)

$$\hat{K}(\alpha) = \hat{K}(x, E; \alpha) \stackrel{\text{def}}{=} \int_{\mathbb{R}^d} e^{(\alpha, s)} K(x, E \times ds), \quad \alpha \in \mathbb{R}^d, x \in \mathbb{S}, E \in \mathcal{E},$$

$$(\lambda_K(\alpha))^{-1} = \text{the convergence parameter of } \hat{K}(\alpha), \quad \text{and} \quad \Lambda_K(\alpha) = \log \lambda_K(\alpha).$$

Let $\{T_i\}_{i \geq 0}$ and $\{(\xi'_i, \xi''_i)\}_{i \geq 0}$ be given as in Lemma 2.1, and let

$$\begin{aligned} \tau &\stackrel{d}{=} T_{i+1} - T_i, & S_\tau &\stackrel{d}{=} (\xi_{T_{i+1}} + \cdots + \xi_{T_{i+1}-1}) + \xi''_{T_i} + \xi'_{T_{i+1}}, \\ \psi(\alpha, \zeta) &= \mathbf{E}_\nu[e^{\langle \alpha, S_\tau \rangle - \zeta \tau}] \quad \text{for all } \alpha \in \mathbb{R}^d, \zeta \in \mathbb{R}, \\ \mathcal{U}_r &= \{\alpha : \psi(\alpha, \zeta) = 1, \text{ some } \zeta < \infty\}. \end{aligned}$$

Observe that $(\mathfrak{M}) \Rightarrow \hat{h}(\alpha) \otimes \hat{\nu}(\alpha) \leq \hat{\mathcal{P}}(\alpha)$, where (for any function h and any measure ν)

$$\hat{h}(x; \alpha) = \int_{\mathbb{R}^d} e^{\langle \alpha, s \rangle} h(x, ds), \quad \hat{\nu}(E; \alpha) = \int_{\mathbb{R}^d} e^{\langle \alpha, s \rangle} \nu(E \times ds).$$

Thus, under (\mathfrak{M}) , the theory for nonnegative kernels may be applied to $\hat{\mathcal{P}}(\alpha)$. This leads to certain representation formulas and other regularity properties for the relevant eigenvectors and eigenvalues, which we now describe.

LEMMA 2.2. *Let $\{(X_n, S_n) : n = 0, 1, \dots\}$ be a Markov additive process satisfying (\mathfrak{M}) .*

(i) *If $\alpha \in \mathcal{U}_r$, then $\hat{\mathcal{P}}(\alpha)$ is $(\lambda_{\mathcal{P}}(\alpha))^{-1}$ -recurrent. Moreover, the eigenvalue $\lambda_{\mathcal{P}}(\alpha)$ and invariant function $r_{\mathcal{P}}(\alpha)$ satisfy the following representation formulas:*

$$(2.3) \quad \psi(\alpha, \Lambda_{\mathcal{P}}(\alpha)) = 1, \quad r_{\mathcal{P}}(x; \alpha) = \mathbf{E}_x[e^{\langle \alpha, S_{T_0} \rangle - \Lambda_{\mathcal{P}}(\alpha) T_0}].$$

(ii) *If $\text{dom } \psi$ is open, then $\text{dom } \Lambda_{\mathcal{P}}$ is open, and on $\text{dom } \Lambda_{\mathcal{P}}$ we have $\alpha \in \mathcal{U}_r$ and $\Lambda_{\mathcal{P}}(\cdot)$ analytic, and $r_{\mathcal{P}}(x; \cdot)$ is finite and analytic on a set $\mathbb{F} \subset \mathbb{S}$, where $\varphi(\mathbb{F}^c) = 0$.*

(iii) *If (\mathfrak{R}) holds and $\alpha \in \text{dom } \Lambda_{\mathcal{P}}$, then $\lambda_{\mathcal{P}}(\alpha)$ is an eigenvalue of $\hat{\mathcal{P}}(\alpha)$, and the associated invariant function $r_{\mathcal{P}}(\alpha)$ is uniformly positive and bounded on $\text{dom } \Lambda_{\mathcal{P}}$ (in particular, if $\hat{\nu}(\alpha)r_{\mathcal{P}}(\alpha) = 1$, then $a \leq \lambda_{\mathcal{P}}(\alpha)r_{\mathcal{P}}(x; \alpha) \leq b$).*

For the proofs, see Ney and Nummelin [(1987a), Sections 3 and 4], and Iscoe, Ney and Nummelin [(1985), Lemma 3.1].

REMARK 2.2. Using the split-chain construction described in Ney and Nummelin [(1984), page 7], the quantities $\Lambda_{\mathcal{P}}(\alpha)$ and $r_{\mathcal{P}}(\cdot; \alpha)$ can be evaluated from (2.3) using direct simulation.

REMARK 2.3. If the lower bound of (\mathfrak{R}) holds and $r_{\mathcal{P}}(\alpha)$ is a ρ -subinvariant function for $\hat{\mathcal{P}}(\alpha)$, then $r_{\mathcal{P}}(\alpha) \geq \rho a(\hat{\nu}(\alpha)r_{\mathcal{P}}(\alpha))$. This implies that $r_{\mathcal{P}}(\alpha)$ is uniformly positive.

Let $\mathcal{P}(x, \cdot) \ll \mathcal{Q}(x, \cdot)$, for all $x \in \mathbb{S}$, and define

$$(2.4) \quad \mathcal{K}_{\mathcal{Q}}(x, dy \times ds) = \left(\frac{d\mathcal{P}}{d\mathcal{Q}}(x, y \times s) \right)^2 \mathcal{Q}(x, dy \times ds).$$

LEMMA 2.3. Assume (\mathfrak{M}) . Then the following hold:

- (i) $(\mathcal{P}(x, E \times \Gamma))^2 \leq \mathcal{K}_{\mathcal{Q}}(x, E \times \Gamma)$, for all $x \in \mathbb{S}$, $E \in \mathfrak{E}$ and $\Gamma \in \mathcal{R}^d$.
- (ii) $(\lambda_{\mathcal{P}}(\alpha))^2 \leq \lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)$, for all $\alpha \in \mathbb{R}^d$. Moreover, if $\alpha \in \mathcal{U}_r$ and $(\lambda_{\mathcal{P}}(\alpha))^2 = \lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)$, then

$$(2.5) \quad \mathcal{Q}(x, dy \times ds) = e^{(\alpha, s) - \Lambda_{\mathcal{P}}(\alpha)} \frac{r_{\mathcal{P}}(y; \alpha)}{r_{\mathcal{P}}(x; \alpha)} \mathcal{P}(x, dy \times ds), \quad \mathcal{P}\text{-a.e. } (y, s),$$

for φ a.e. x , where $r_{\mathcal{P}}(\alpha)$ is the $(\lambda_{\mathcal{P}}(\alpha))^{-1}$ -invariant function for $\hat{\mathcal{P}}(\alpha)$. Conversely, if $\alpha \in \mathcal{U}_r$ and \mathcal{Q} is defined by (2.5), for all $x \in \mathbb{S}$, then $(\lambda_{\mathcal{P}}(\alpha))^2 = \lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)$.

(iii) If \mathcal{Q} is defined as in (2.5) and $\alpha, \beta \in \mathcal{U}_r$, then $\lambda_{\mathcal{K}_{\mathcal{Q}}}(\alpha + \beta) = \lambda_{\mathcal{P}}(\alpha)\lambda_{\mathcal{P}}(\beta)$, and the associated invariant functions satisfy the equation $r_{\mathcal{K}_{\mathcal{Q}}}(\alpha + \beta) = r_{\mathcal{P}}(\alpha)r_{\mathcal{P}}(\beta)$.

PROOF. Part (i) is established using Hölder's inequality.

For (ii), assume $\lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha) < \infty$, and let $r_{\mathcal{K}_{\mathcal{Q}}}$ be a $(\lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha))^{-1}$ -subinvariant function for $\hat{\mathcal{K}}_{\mathcal{Q}}(2\alpha)$. Apply Hölder's inequality to the integral

$$\int_{\mathbb{S} \times \mathbb{R}^d} e^{(\alpha, s)} r_{\mathcal{K}_{\mathcal{Q}}}(y; 2\alpha)^{1/2} \frac{d\mathcal{P}}{d\mathcal{Q}}(x, y \times s) \mathcal{Q}(x, dy \times ds)$$

to obtain

$$(2.6) \quad \hat{\mathcal{P}}(\alpha) r_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)^{1/2} \leq \lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)^{1/2} r_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)^{1/2}.$$

Thus $r_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)^{1/2}$ is a $(\lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha))^{-1/2}$ -subinvariant function for $\hat{\mathcal{P}}(\alpha)$. Hence $(\lambda_{\mathcal{P}}(\alpha))^2 \leq \lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)$ [Nummelin (1984), Proposition 5.2].

Now suppose $\alpha \in \mathcal{U}_r$ and $(\lambda_{\mathcal{P}}(\alpha))^2 = \lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)$. Then by (2.6), $r_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)^{1/2}$ is a $(\lambda_{\mathcal{P}}(\alpha))^{-1}$ -subinvariant function for $\hat{\mathcal{P}}(\alpha)$. It follows that $r_{\mathcal{P}}(\alpha) = C(r_{\mathcal{K}_{\mathcal{Q}}}(2\alpha))^{1/2} \varphi$ a.e., for some positive constant C [Nummelin (1984), Theorem 5.1]. Then there is equality in (2.6), namely equality in Hölder's inequality, and—after normalizing so that \mathcal{Q} is a probability measure—this implies (2.5).

Conversely, note that (2.5) implies

$$(2.7) \quad \mathcal{K}_{\mathcal{Q}}(x, E \times \Gamma) = \int_{E \times \Gamma} \lambda_{\mathcal{P}}(\alpha) e^{-(\alpha, s)} \frac{r_{\mathcal{P}}(x; \alpha)}{r_{\mathcal{P}}(y; \alpha)} \mathcal{P}(x, dy \times ds)$$

for all $E \in \mathfrak{E}$, $\Gamma \in \mathcal{R}^d$,

and hence

$$(2.8) \quad \hat{\mathcal{K}}_{\mathcal{Q}}(2\alpha) (r_{\mathcal{P}}(\alpha))^2 = (\lambda_{\mathcal{P}}(\alpha))^2 (r_{\mathcal{P}}(\alpha))^2.$$

It follows that $(\lambda_{\mathcal{P}}(\alpha))^2 = \lambda_{\mathcal{K}_{\mathcal{Q}}}(2\alpha)$.

To establish (iii), repeat (2.7) and (2.8) with “ $r_{\mathcal{P}}(\alpha)r_{\mathcal{P}}(\beta)$ ” in place of “ $r_{\mathcal{P}}^2(\alpha)$.” \square

3. Main results.

3.1. *Notation, hypotheses and estimation results.* Given a Markov additive process $\{(X_n, S_n) : n = 0, 1, \dots\}$, we would like to evaluate $\mathbf{P}\{T^\varepsilon(A) < \infty\}$, where $T^\varepsilon(A)$ is defined as in (1.2). Suppose that we simulate for this quantity using simulated random variables $\{(\tilde{X}_n^\varepsilon, \tilde{S}_n^\varepsilon) : n = 0, 1, \dots\}$ with transition kernel

$$\mathcal{Q}^{n,\varepsilon}(x, E \times \Gamma) = \mathbf{P}\{(\tilde{X}_{n+1}^\varepsilon, \tilde{S}_{n+1}^\varepsilon) \in E \times \Gamma \mid \tilde{X}_n^\varepsilon = x\}.$$

If $\mathcal{P}(x, \cdot) \ll \mathcal{Q}^{n,\varepsilon}(x, \cdot)$, for all $n \in \mathbb{Z}_+$, $\varepsilon > 0$ and $x \in \mathbb{S}$, then

$$(3.1) \quad \begin{aligned} & \mathbf{P}\{T^\varepsilon(A) < \infty\} \\ &= \sum_k \int_{\mathfrak{P}_k^\varepsilon} \left(\prod_{n=0}^{k-1} \frac{d\mathcal{P}}{d\mathcal{Q}^{n,\varepsilon}}(x_n, x_{n+1} \times s_{n+1}) \right) \\ & \quad \times \mathcal{Q}^{0,\varepsilon}(x_0, dx_1 \times ds_1) \cdots \mathcal{Q}^{k-1,\varepsilon}(x_{k-1}, dx_k \times ds_k), \end{aligned}$$

where $\mathfrak{P}_k^\varepsilon$ denotes all paths which *first* hit A/ε at time k , that is,

$$(3.2) \quad \mathfrak{P}_k^\varepsilon = \left\{ (x_0, \dots, x_k; s_0, \dots, s_k) : \sum_{j=1}^l s_j \in \frac{A}{\varepsilon} \text{ for } l = k \text{ but not for } l < k \right\}.$$

It follows from (3.1) that

$$(3.3) \quad \mathcal{E}_{\mathcal{Q},\varepsilon} \stackrel{\text{def}}{=} \left(\prod_{n=0}^{T^\varepsilon(A)-1} \frac{d\mathcal{P}}{d\mathcal{Q}^{n,\varepsilon}}(\tilde{X}_n^\varepsilon, \tilde{X}_{n+1}^\varepsilon \times \tilde{S}_{n+1}^\varepsilon) \right) \mathbb{1}_{\{T^\varepsilon(A) < \infty\}}$$

is an unbiased estimator for $\mathbf{P}\{T^\varepsilon(A) < \infty\}$.

The efficiency of this estimator is measured by its variance, which we will study in an asymptotic sense as $\varepsilon \rightarrow 0$. Since $\text{Var}(\mathcal{E}_{\mathcal{Q},\varepsilon}) = \mathbf{E}(\mathcal{E}_{\mathcal{Q},\varepsilon}^2) - (\mathbf{E}(\mathcal{E}_{\mathcal{Q},\varepsilon}))^2$, and

$$\mathbf{E}(\mathcal{E}_{\mathcal{Q},\varepsilon}) = \mathbf{P}\{T^\varepsilon(A) < \infty\}$$

has the asymptotic characterization given in (1.4) [Collamore (1996a), Theorems 2.1 and 2.2], it is sufficient to study the asymptotic behavior of

$$(3.4) \quad \mathbf{E}(\mathcal{E}_{\mathcal{Q},\varepsilon}^2) = \sum_k \int_{\mathfrak{P}_k^\varepsilon} \mathcal{K}_{\mathcal{Q}}^{0,\varepsilon}(x_0, dx_1 \times ds_1) \cdots \mathcal{K}_{\mathcal{Q}}^{k-1,\varepsilon}(x_{k-1}, dx_k \times ds_k),$$

where, for all n, ε ,

$$(3.5) \quad \begin{aligned} & \mathcal{K}_{\mathcal{Q}}^{n,\varepsilon}(x_n, dx_{n+1} \times ds_{n+1}) \\ & \stackrel{\text{def}}{=} \left(\frac{d\mathcal{P}}{d\mathcal{Q}^{n,\varepsilon}}(x_n, x_{n+1} \times s_{n+1}) \right)^2 \mathcal{Q}^{n,\varepsilon}(x_n, dx_{n+1} \times ds_{n+1}). \end{aligned}$$

Our objective will be to give estimation results for $\mathbf{E}(\mathcal{E}_{\mathcal{Q},\varepsilon}^2)$ as $\varepsilon \rightarrow 0$, and optimality results describing *which* transition kernels \mathcal{Q} for the simulated random variables minimize $\mathbf{E}(\mathcal{E}_{\mathcal{Q},\varepsilon}^2)$ as $\varepsilon \rightarrow 0$.

We first introduce some additional notation and hypotheses, as follows. Let

$$\begin{aligned} \text{cone}(C) &= \{\zeta v : \zeta \geq 0, v \in C\} \text{ for any } C \subset \mathbb{R}^d, \\ \text{cone}_\delta(C) &= \{\zeta v : \zeta \geq 0, \|v - w\| < \delta \|w\|, \text{ some } w \in C\} \text{ for any } \delta > 0, \end{aligned}$$

$$\mathfrak{S} = \text{cone}\left(\text{Supp}_v \frac{S_\tau}{\tau}\right) \text{ with } v, \tau \text{ as in Section 2,}$$

$$C^\perp = \{v : \langle \alpha, v \rangle \leq 0, \text{ all } \alpha \in C\} \text{ for any } C \subset \mathbb{R}^d,$$

$$\mathcal{H}(\alpha, a) = \{v : \langle \alpha, v \rangle > a\} \text{ for any } \alpha \in \mathbb{R}^d \text{ and } a \in \mathbb{R},$$

$$\mathcal{L}_a f = \{v : f(v) \leq a\} \text{ for any } f: \mathbb{R}^d \rightarrow \mathbb{R} \text{ and } a \in \mathbb{R},$$

$$\text{dom } f = \{v : f(v) < \infty\} \text{ for any } f: \mathbb{R}^d \rightarrow \mathbb{R}.$$

For any nonnegative φ -irreducible kernel K , let $\hat{K}^0(\alpha) = I$, $\hat{K}^n(\alpha) = \hat{K}(\alpha)\hat{K}^{n-1}(\alpha)$, $n = 1, 2, \dots$, and

$$\bar{\Lambda}_K(\alpha) = \limsup_{n \rightarrow \infty} n^{-1} \log \hat{K}^n(X_0, \mathbb{S}; \alpha),$$

$$\Lambda_K^{(N)}(\alpha) = \sup_{n \geq N} n^{-1} \log \hat{K}^n(X_0, \mathbb{S}; \alpha),$$

$$I_K(v) = \sup\{\langle \alpha, v \rangle : \alpha \in \mathcal{L}_0 \Lambda_K\}, \quad \mathfrak{D}_K = \text{dom } I_K,$$

$$I_K^{(c)}(v) = \sup\{\langle \alpha, v \rangle : \alpha \in \mathcal{L}_c \Lambda_K\}, \quad \mathfrak{D}_K^{(c)} = \text{dom } I_K^{(c)},$$

$$\bar{I}_K(v) = \sup\{\langle \alpha, v \rangle : \alpha \in \mathcal{L}_0 \bar{\Lambda}_K\}, \quad \text{respectively } \bar{I}_K^{(c)}(\cdot);$$

where, as in the previous section, $(\lambda_K(\alpha))^{-1}$ is the convergence parameter of $\hat{K}(\alpha)$ and $\Lambda_K(\alpha) = \log \lambda_K(\alpha)$. In the definitions of I_K , $I_K^{(c)}$ and \bar{I}_K , we follow the convention that the supremum over the empty set equals $-\infty$.

For any set C , let $\mathbb{1}_C(v)$ denote the indicator function of C (1 for $v \in C$; 0 for $v \notin C$); let $\text{ri } C$ denote the relative interior of C ; and let ∂C denote the relative boundary of C . For any function f , let f^* denote the convex conjugate of f . [For definitions, see Rockafellar (1970).]

Next we turn to certain regularity conditions on the Markov additive process $\{(X_n, S_n) : n = 0, 1, \dots\}$ and the set A which we will often need to impose. If, for example, S_1, S_2, \dots are the sums of an i.i.d. sequence of random variables, or the additive sums of a Markov additive process which satisfies (\mathfrak{A}) under both \mathcal{P} and $\mathcal{Q}^{n,\varepsilon} = \mathcal{Q}$, then

$$\Lambda_{\mathcal{K}_\mathcal{Q}}^{(1)}(\alpha) < \infty \text{ for all } \alpha \in \text{dom } \bar{\Lambda}_{\mathcal{K}_\mathcal{Q}}.$$

For general Markov additive processes, this property need not necessarily hold, but we will often assume that the following weaker condition is satisfied.

HYPOTHESIS.

(H1) $\Lambda_{\mathcal{K}_Q}^{(1)}(\alpha) < \infty$, for all $\alpha \in \mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_Q}$.

A second assumption which we would like to impose is that the mean drift of the process is directed *away* from the set A . Now, under natural conditions,

$$(3.6) \quad \Lambda_{\mathcal{P}}^*(v) = 0 \quad \Leftrightarrow \quad \mathbf{E}_\pi(S_1) = v,$$

where π is the stationary measure of the Markov chain $\{X_n\}$; this is true, for example, if (\mathfrak{R}) holds and $0 \in \text{dom } \Lambda_{\mathcal{P}}$, or alternatively if (\mathfrak{M}') holds and the set $\text{dom } \psi$ is open [Ney and Nummelin (1987a), Lemma 3.3 and Lemma 5.2]. From (3.6), we then see that the central tendency of S_1, S_2, \dots is along the mean ray $\{\zeta v : \zeta \geq 0 \text{ and } v \in \mathcal{L}_0 \Lambda_{\mathcal{P}}^*\}$. In fact, we will assume slightly more, namely that the set A is disjoint from a small δ -cone about this mean ray, or more formally:

(H2) $\text{cl } A \cap \text{cone}_\delta(\mathcal{L}_0 \Lambda_{\mathcal{P}}^*) = \emptyset$, for some $\delta > 0$.

DEFINITION. We say that a family of probability measures $\{\mathcal{Q}^{n,\varepsilon}(x, E \times \Gamma) : E \in \mathfrak{E}, \Gamma \in \mathcal{R}^d\}$ belongs to the class \mathcal{C}_0 if $\mathcal{Q}^{n,\varepsilon}(x, \cdot) = \mathcal{Q}(x, \cdot)$ for all $x \in \mathbb{S}$, independent of n and ε , and $\mathcal{P}(x, \cdot) \ll \mathcal{Q}(x, \cdot)$, for all $x \in \mathbb{S}$.

DEFINITION. If $A \subset \mathbb{R}^d$, then we say that $v \in \text{cl } A - \{0\}$ is an *exposed point* of A if the line segment $\{\zeta v : 0 \leq \zeta < 1\}$ does not intersect $\text{cl } A$.

THEOREM 3.1. *Let A be a convex open set intersecting $\text{ri } \mathfrak{S}$, $0 \notin \text{int } A$. Let \mathfrak{A} denote the exposed points of A . Assume that $\{(X_n, S_n) : n = 0, 1, \dots\}$ satisfies (\mathfrak{M}') and has initial state $X_0 = x_0$. Suppose that simulation is performed with a kernel $\mathcal{Q} \in \mathcal{C}_0$. Then the following hold:*

(i) *Lower bound.*

$$(3.7) \quad \liminf_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2) \geq - \inf_{v \in \mathfrak{A}} I_{\mathcal{K}_Q}(v).$$

(ii) *Upper bound. Assume, in addition, that the following holds:*

(C) $\inf_\alpha \bar{\Lambda}_{\mathcal{K}_Q}(\alpha) < 0$, (H1) is satisfied and $A \cap \text{cone}_\delta(\mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_Q})^\perp = \emptyset$, for some $\delta > 0$.

Then, for φ a.e. x_0 ,

$$(3.8) \quad \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2) \leq - \inf_{v \in \mathfrak{A}} \bar{I}_{\mathcal{K}_Q}(v).$$

If the lower bound of (\mathfrak{R}) is satisfied, then in place of (C) it is sufficient to assume $A \cap (\mathcal{L}_0 \Lambda_{\mathcal{K}_Q})^\perp = \emptyset$; and then (3.8) holds with $I_{\mathcal{K}_Q}$ in place of $\bar{I}_{\mathcal{K}_Q}$.

REMARK 3.1. (i) Some conditions under which $I_{\mathcal{K}_Q} = \bar{I}_{\mathcal{K}_Q}$ are described in the discussion following Lemma 3.2. However, it is known in the context of large deviations for Markov additive processes that these bounds need *not* be the same in general; see de Acosta and Ney [(1998), Section 4].

(ii) Since $(\mathcal{L}_0 \Lambda_{\mathcal{K}_Q})^\perp = \{v : I_{\mathcal{K}_Q}(v) \leq 0\}$, “ $A \cap (\mathcal{L}_0 \Lambda_{\mathcal{K}_Q})^\perp = \emptyset$ ” holds as long as $\inf\{I_{\mathcal{K}_Q}(v) : v \in \mathfrak{A}\} > 0$. The weakening of this assumption to the case where $\mathbf{E}(\mathcal{E}_{Q,\varepsilon}^2)$ exhibits exponential growth as $\varepsilon \rightarrow 0$ is apparently not possible, in general.

The stronger condition “ $A \cap \text{cone}_\delta(\mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_Q})^\perp = \emptyset$, for some $\delta > 0$ ” is similar to, and in fact a strengthening of, condition (H2) (for $\mathcal{O} \notin \text{cl} A$). In particular, $2(\mathcal{L}_0 \Lambda_{\mathcal{P}}) \supset \mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_Q}$ [Lemma 2.3(ii), since $\bar{\Lambda}_{\mathcal{K}_Q} \geq \Lambda_{\mathcal{K}_Q}$]. It follows that $\mathcal{L}_0 \bar{I}_{\mathcal{K}_Q} \supset \mathcal{L}_0 I_{\mathcal{P}} = \{\zeta v : \zeta \geq 0, v \in \mathcal{L}_0 \Lambda_{\mathcal{P}}^*\}$.

REMARK 3.2. If the lower bound of (\mathfrak{A}) holds, then as an upper bound we actually obtain

$$(3.9) \quad \mathbf{E}(\mathcal{E}_{Q,\varepsilon}^2) \leq \text{const} \cdot \exp\left\{-\varepsilon^{-1} \inf_{v \in \mathfrak{A}} I_{\mathcal{K}_Q}(v)\right\}.$$

REMARK 3.3. If $S_n = \xi_1 + \dots + \xi_n$, where $\{\xi_n\}$ is an i.i.d. sequence of random variables, then the quantities $\Lambda_{\mathcal{P}}$ and $\Lambda_{\mathcal{K}_Q}$, which determine the rate functions on the right-hand sides of (3.7) and (3.8), can be simplified. In this setting, $\Lambda_{\mathcal{P}} = \bar{\Lambda}_{\mathcal{P}}$ may be identified as the cumulant generating function of ξ_n , namely,

$$\Lambda_{\mathcal{P}}(\alpha) = \log \int_{\mathbb{R}^d} e^{\langle \alpha, s \rangle} \mathcal{P}(ds),$$

where \mathcal{P} is the probability law of ξ_n , and similarly for $\Lambda_{\mathcal{K}_Q}$. Furthermore, any discussion of invariant functions may be dropped; that is, we may always take $r_{\mathcal{P}}(\cdot; \alpha) = 1$ and $r_{\mathcal{K}_Q}(\cdot; \alpha) = 1$.

EXAMPLE 3.1. Let $S_n = \xi_1 + \dots + \xi_n$, where $\{\xi_n\}_{n \in \mathbb{Z}_+} \subset \mathbb{R}^2$ is an i.i.d. sequence of normal random variables with mean $m = (\mu, \mu)$ and covariance $S = \begin{pmatrix} 1 & \sigma \\ \sigma & 1 \end{pmatrix}$, where $\mu > 0$ and $-1 < \sigma < 1$. Let $A = \{(v_1, v_2) : v_1 < -1 \text{ and } v_2 < -1\}$. We consider the simulation of $\mathbf{P}\{T^\varepsilon(A) < \infty\}$ using an exponentially tilted distribution of the form

$$(3.10) \quad \mathcal{Q}_\beta(ds) = e^{\langle \beta, s \rangle - \Lambda_{\mathcal{P}}(\beta)} \mathcal{P}(ds).$$

By Lemma 2.3(iii), $\alpha \in \mathcal{L}_0 \Lambda_{\mathcal{K}_Q} \Leftrightarrow \Lambda_{\mathcal{P}}(\beta) + \Lambda_{\mathcal{P}}(\alpha - \beta) \leq 0$. Since the cumulant generating function for a Normal(m, S) random variable is $\Lambda_{\mathcal{P}}(\alpha) = \langle \alpha, m \rangle + \frac{1}{2} \langle \alpha, S \alpha \rangle$, it follows from a straightforward computation that

$$(3.11) \quad \mathcal{L}_0 \Lambda_{\mathcal{K}_{Q_{\bar{\beta}}}} = \left\{ \bar{\alpha} : (1 + \sigma) \left(\bar{\alpha}_1 + \frac{\sqrt{2}\mu}{1 + \sigma} - \bar{\beta}_1 \right)^2 + (1 - \sigma) (\bar{\alpha}_2 - \bar{\beta}_2)^2 \leq b \right\},$$

where $b = -(1 + \sigma)\bar{\beta}_1^2 - 2\sqrt{2}\mu\bar{\beta}_1 - (1 - \sigma)\bar{\beta}_2^2 + 2\mu^2/(1 + \sigma)$, and $\bar{\alpha}, \bar{\beta}$ denote the values of α, β in a coordinate system which has been rotated by angle $\pi/4$.

Our objective is to apply Theorem 3.1 to analyze the dependence of $\mathbf{E}(\mathcal{E}_{\mathcal{K}_{\bar{\beta}}}^2)$ on $\bar{\beta}$. Thus we would like to study

$$(3.12) \quad J(\bar{\beta}) = - \inf_{v \in \partial A} I_{\mathcal{K}_{\bar{\beta}}}(v)$$

as a function of $\bar{\beta}$. [If $J \geq 0$, the right-hand side of (3.8) must be taken to be infinity.]

Suppose for simplicity that $\mu = 1/\sqrt{2}$ and $\sigma = 1/2$. The function $J(\bar{\beta})$ can then be analytically computed from (3.11) for all values of $\bar{\beta}$. For example, if r is a sufficiently large positive constant, then the level sets where $J(\bar{\beta}) = -r < 0$ are given by

$$(3.13) \quad \left(\bar{\beta}_1 + \frac{r}{2\sqrt{2}}\right)^2 + \frac{1}{6}\bar{\beta}_2^2 = r\left(\frac{\sqrt{2}}{3} - \frac{r}{8}\right).$$

A graph of the level sets of J over all of \mathbb{R}^2 is given in Figure 1.

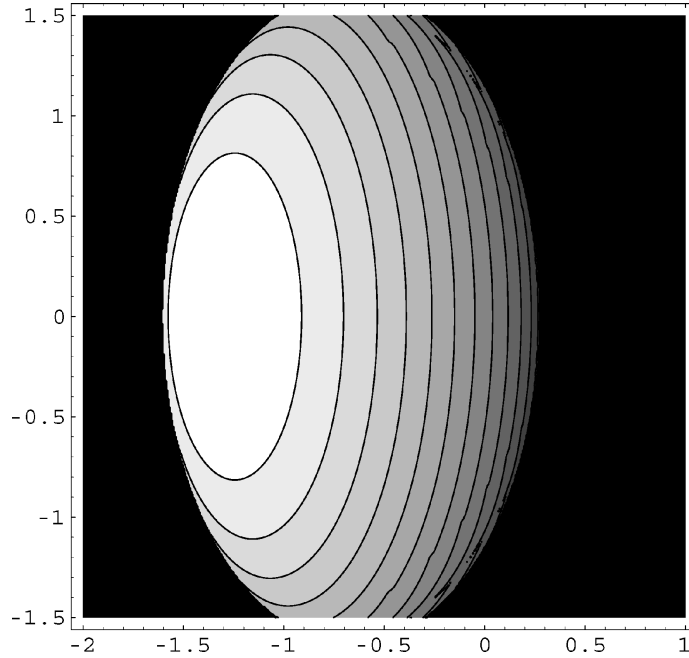


FIG. 1. Let $a = \inf\{J(\bar{\beta}) : \bar{\beta} \in \mathbb{R}^2\}$, where J is defined as in (3.12). The figure illustrates the level lines $r = a + 0.25, a + 0.5, a + 0.75, \dots$, with $\bar{\beta}_1$ on the horizontal axis, and $\bar{\beta}_2$ on the vertical axis; J is seen to increase rapidly to the left of its minimum at $(-4/3, 0)$. The black area indicates the region where $J = \infty$.

The minimum value of J occurs at the maximum r for which the right-hand side of (3.13) ≥ 0 , that is, $r = 8\sqrt{2}/3$, and for this r we obtain by (3.13) that $\bar{\beta} = (-4/3, 0)$. At the other extreme, the points where $J = \infty$ are all contained in the complement of the zero-set $\mathcal{L}_0\Lambda_{\mathcal{P}} = \{\bar{\alpha} : 3(\bar{\alpha}_1 + 2/3)^2 + \bar{\alpha}_2^2 \leq 4/3\}$. This illustrates the general fact that $J(\bar{\beta})$ tends to be smaller on $\mathcal{L}_0\Lambda_{\mathcal{P}}$ as compared with $(\mathcal{L}_0\Lambda_{\mathcal{P}})^c$.

EXAMPLE 3.2. Let $\{Y_n\}_{n \in \mathbb{Z}_+}$ be an ARMA(p, q) process in \mathbb{R}^d , namely,

$$Y_n = -(\phi_1 Y_{n-1} + \dots + \phi_p Y_{n-p}) + W_n + \theta_1 W_{n-1} + \dots + \theta_q W_{n-q}$$

for constants $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ satisfying appropriate regularity conditions, as given in Brockwell and Davis [(1991), Chapter 3]. For simplicity, take $\{W_n\}_{n \in \mathbb{Z}_+}$ to be i.i.d. Normal(0, S). As in Meyn and Tweedie [(1993), page 28], we may then write $Y_n = F(X_n)$, where $\{X_n\}$ is a Markov chain taking values in \mathbb{R}^{ld} , $l = \max\{p, q + 1\}$, and this Markov chain can be shown to satisfy (\mathfrak{M}') [take $h(x) = \text{const} \cdot \mathbb{1}_{\mathcal{O}_\varepsilon}(x)$, where $\mathcal{O}_\varepsilon = [-\varepsilon, \varepsilon]^l$].

Assume that the past history of the process is known or, equivalently, that the initial state of the Markov chain is $X_0 = x_0$, for some $x_0 \in \mathbb{R}^{ld}$. Let $m \in \mathbb{R}^d - \{0\}$ and $\xi_n = Y_n + m$; and let $S_n = \xi_1 + \dots + \xi_n$, $n \geq 1$, and $S_0 = 0$. Then $\{(X_n, S_n) : n = 0, 1, \dots\}$ is a Markov additive process.

A simple computation gives

$$(3.14) \quad \bar{\Lambda}_{\mathcal{P}}(\alpha) = \langle \alpha, m \rangle + \sum_{j=0}^{\infty} \Psi_j \langle \alpha, S\alpha \rangle$$

for certain constants $\{\Psi_j\}$ [cf. Brockwell and Davis (1991), Theorem 3.1.1].

Next we observe that actually $\Lambda_{\mathcal{P}} = \bar{\Lambda}_{\mathcal{P}}$. To this end, note that

$$(3.15) \quad \begin{aligned} \Lambda_{\mathcal{P}}(\alpha) &= \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{E}[e^{\langle \alpha, S_n \rangle} \mathbb{1}_{\mathcal{O}_\varepsilon}(X_n)] \\ &= \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{E}[e^{\langle \alpha, S_n \rangle} \mathbb{1}_{A_{bn}}(X_{\lfloor bn \rfloor}) \mathbb{1}_{\mathcal{O}_\varepsilon}(X_n)], \end{aligned}$$

where $b \in (0, 1)$, $A_k = [-\sqrt{ka}, \sqrt{ka}]^p$, and a is a suitably large positive constant; the last step was obtained by an application of Hölder's inequality to $(e^{\langle \alpha, S_n \rangle} \mathbb{1}_{\mathcal{O}_\varepsilon}(X_n))(\mathbb{1}_{A_{bn}^c}(X_{\lfloor bn \rfloor}))$. Moreover,

$$(3.16) \quad \frac{1}{n} \log \mathbf{E}[e^{\langle \alpha, S_n - S_{\lfloor bn \rfloor} \rangle} \mathbb{1}_{\mathcal{O}_\varepsilon}(X_n) \mid X_{\lfloor bn \rfloor} = x] = (1 - b)\Lambda_{\mathcal{P},n}(\alpha) + \Delta_n(x),$$

where $\Delta_n(x) \rightarrow 0$ as $n \rightarrow \infty$, uniformly for $x \in A_{bn}$, and $\Lambda_{\mathcal{P},n}(\alpha) \rightarrow \Lambda_{\mathcal{P}}(\alpha)$. Substitute (3.16) into (3.15), let $n \rightarrow \infty$ and then $b \rightarrow 1$ and apply Hölder's inequality once more to obtain $\Lambda_{\mathcal{P}}(\alpha) = \bar{\Lambda}_{\mathcal{P}}(\alpha)$, for all α .

Also, by a direct computation, $r(x; \alpha) = \exp\{c, x\}$ for some $c \in \mathbb{R}^l$.

Suppose that we simulate using an exponentially tilted distribution of the form (2.5) (with “ β ” in place of “ α ”). A repetition of the above argument yields $\bar{\Lambda}_{\mathcal{K}_{\mathcal{Q}_\beta}}(\alpha) = \Lambda_{\mathcal{K}_{\mathcal{Q}_\beta}}(\alpha) = \Lambda_{\mathcal{P}}(\beta) + \Lambda_{\mathcal{P}}(\alpha - \beta)$, by Lemma 2.3(iii). We may now proceed as in the previous example to determine the interdependence of $\mathbf{E}(\mathcal{E}_{\mathcal{Q}_\beta, \varepsilon}^2)$ and \mathcal{Q}_β .

For further applications to Markov and semi-Markov processes, see, for example, Iscoe, Ney and Nummelin (1985), Ney and Nummelin (1987a) and Meyn and Tweedie (1993).

3.2. *Optimality.* Our next objective is to find an optimal $\mathcal{Q} \in \mathcal{C}_0$ which maximizes the decay rate on the right-hand sides of (3.7) and (3.8). For this purpose, first recall from (1.4) that the decay of $\mathbf{P}\{T^\varepsilon(A) < \infty\}$ is governed by the rate function $I_{\mathcal{P}}$.

To obtain an optimal simulation regime for the multidimensional problem, a very essential role will be played by the following.

LEMMA 3.2. *Let $A \subset \mathbb{R}^d$ be a convex set intersecting $\text{ri } \mathfrak{S}$. Suppose that the probability law of S_n/n satisfies the large deviation principle with rate function $I_{\mathcal{P}} = \Lambda_{\mathcal{P}}^*$. Assume that $\text{dom } \Lambda_{\mathcal{P}}$ is open and (H2) is satisfied. Let $a = \inf_{v \in A} I_{\mathcal{P}}(v)$. Then the following hold:*

- (i) *an element $\alpha_0 \in \partial(\mathcal{L}_0 \Lambda_{\mathcal{P}})$ determines a hyperplane which separates A and $\mathcal{L}_a I_{\mathcal{P}}$, with $A \subset \{v : \langle \alpha_0, v \rangle \geq a\}$ and $\mathcal{L}_a I_{\mathcal{P}} \subset \{v : \langle \alpha_0, v \rangle \leq a\}$;*
- (ii) *there exists a unique element $v_0 \in \text{cl } A$ such that $I_{\mathcal{P}}(v_0) = a$;*
- (iii) *$\inf_{v \in A} I_{\mathcal{P}}(v) = \inf_{v \in \partial A} I_{\mathcal{P}}(v) = \langle \alpha_0, v_0 \rangle$;*
- (iv) *if $\text{int } \mathfrak{S} \neq \emptyset$, then α_0 is the unique element of the subgradient set $\partial I_{\mathcal{P}}(v_0)$;*
- (v) *the gradient of $\Lambda_{\mathcal{P}}$ at α_0 points in the same direction as v_0 , that is, $v_0 = \varrho \nabla \Lambda_{\mathcal{P}}(\alpha_0)$ for some constant $\varrho > 0$.*

The proof of Lemma 3.2 is given in Collamore [(1996a), Lemma 3.2; (1998), Lemma 2.2]. The uniqueness in (iv) is obtained from the strict convexity of $\Lambda_{\mathcal{P}}$ [Collamore (1996b), page 38, Ney and Nummelin (1987a), Corollary 3.3]. If $\text{dom } \Lambda_{\mathcal{P}}$ is open and (\mathfrak{M}') is satisfied, then a sufficient condition for the probability law of S_n/n to satisfy the large deviation principle with rate $I_{\mathcal{P}} = \Lambda_{\mathcal{P}}^*$ is that $\Lambda_{\mathcal{P}} = \bar{\Lambda}_{\mathcal{P}}$. [See Ney and Nummelin (1987b). The generating function $\bar{\Lambda}_{\mathcal{P}}$ is the same as that appearing in the Gärtner–Ellis theorem, as given, e.g., in Dembo and Zeitouni (1998), Theorem 2.3.6.]

In our next theorem, we will work with a hypothesis which extends this condition “ $\Lambda_{\mathcal{P}} = \bar{\Lambda}_{\mathcal{P}}$ ” to the kernel $\mathcal{K}_{\mathcal{Q}}$.

HYPOTHESIS.

(H3) $\Lambda_K(\cdot) = \bar{\Lambda}_K(\cdot)$ for $K = \mathcal{P}$ and $K = \mathcal{K}_{\mathcal{Q}}$.

A sufficient condition for (H3) to hold is any one of the following:

(i) S_1, S_2, \dots are the sums of an i.i.d. sequence of random variables, or the additive sums in a Markov additive process $\{(X_n, S_n) : n = 0, 1, \dots\}$ where the state space of $\{X_n\}$ is finite.

(ii) $\{(X_n, S_n) : n = 0, 1, \dots\}$ is a Markov additive process on a general state space, and the lower bound of (\mathfrak{R}) is satisfied.

(iii) $\{(X_n, S_n) : n = 0, 1, \dots\}$ is a Markov additive process on a general state space, and the entire state space is an “ s -set.” [See Ney and Nummelin (1987a, b).]

There are, of course, other situations where (i)–(iii) are difficult or impossible to verify, but (H3) nonetheless holds. For example, this condition was verified directly for ARMA(p, q) processes in Example 3.2. [As with hypothesis (H1), it would actually be enough to assume that $\Lambda_K = \bar{\Lambda}_K$ on $\mathcal{L}_0\Lambda_K$, for $K = \mathcal{P}$ and $K = \mathcal{K}_Q$.]

A second condition which is needed in Lemma 3.2 is that the domain of $\Lambda_{\mathcal{P}}$ is an open set. A sufficient condition for this to hold is that $\text{dom } \psi$ is open [Lemma 2.2(ii)].

Our present objective is to apply Theorem 3.1 and Lemma 3.2 to obtain an optimal simulation distribution which maximizes the decay rate on the right-hand sides of (3.7) and (3.8). To this end, first note by Lemma 2.3(ii) that $\mathcal{L}_0\Lambda_{\mathcal{K}_Q} \subset 2(\mathcal{L}_0\Lambda_{\mathcal{P}})$. Hence $I_{\mathcal{K}_Q}(v) \leq 2I_{\mathcal{P}}(v)$ for all $v \in \mathbb{R}^d$. Now focus on this inequality at the special point $v = v_0$ of Lemma 3.2(ii). By Lemma 3.2(iv) and Rockafellar [(1970), Theorem 23.5]

$$(3.17) \quad \mathcal{L}_0\Lambda_{\mathcal{P}} \cap \{\alpha : \langle \alpha - \alpha_0, v_0 \rangle \geq 0\} = \{\alpha_0\}.$$

Since $\mathcal{L}_0\Lambda_{\mathcal{K}_Q} \subset 2(\mathcal{L}_0\Lambda_{\mathcal{P}})$, it follows from (3.17) that the only way to obtain the equality $I_{\mathcal{K}_Q}(v_0) = 2I_{\mathcal{P}}(v_0)$ ($= 2\langle \alpha_0, v_0 \rangle$) is to have $2\alpha_0 \in \mathcal{L}_0\Lambda_{\mathcal{K}_Q}$. By Lemma 2.3(ii), this occurs precisely when $Q = Q^*$, where

$$(3.18) \quad Q^*(x, dy \times ds) = e^{\langle \alpha_0, s \rangle} \frac{r_{\mathcal{P}}(y; \alpha_0)}{r_{\mathcal{P}}(x; \alpha_0)} \mathcal{P}(x, dy \times ds).$$

Hence

$$(3.19) \quad I_{\mathcal{K}_Q}(v_0) \leq 2I_{\mathcal{P}}(v_0) \quad \text{with equality} \Leftrightarrow Q = Q^* \text{ for } \mathcal{P} \text{ a.e. } (y, s),$$

where Q^* is given by (3.18) [φ a.e. x , in the sense of Lemma 2.3(ii)]. From (3.19) and Lemma 3.2(ii), it follows that if $Q \neq Q^*$, then

$$(3.20) \quad \inf_{v \in A} I_{\mathcal{K}_Q}(v) < 2I_{\mathcal{P}}(v_0) = 2 \inf_{v \in A} I_{\mathcal{P}}(v).$$

Conversely, suppose that $Q = Q^*$. Then by Lemma 2.3(iii), $\mathcal{L}_0\Lambda_{\mathcal{K}_{Q^*}} = \{\alpha + \alpha_0 : \alpha \in \mathcal{L}_0\Lambda_{\mathcal{P}}\}$. Hence

$$(3.21) \quad I_{\mathcal{K}_{Q^*}}(v) = \langle \alpha_0, v \rangle + I_{\mathcal{P}}(v).$$

It follows from Lemma 3.2(i), (iii) that

$$(3.22) \quad 2 \inf_{v \in A} I_{\mathcal{P}}(v) = \inf_{v \in A} I_{\mathcal{K}_{\mathcal{Q}^*}}(v) \quad \left(= \inf_{v \in A} \bar{I}_{\mathcal{K}_{\mathcal{Q}^*}}(v) \text{ by (H3)} \right).$$

Finally, we observe that the hypothesis “ $A \cap \text{cone}_{\delta}(\mathcal{L}_0 \Lambda_{\mathcal{K}_{\mathcal{Q}^*}})^{\perp} = \emptyset$, some $\delta > 0$ ” in the upper bound of Theorem 3.1 is satisfied for $\mathcal{Q} = \mathcal{Q}^*$. To this end, note by (3.21) that $\mathcal{L}_0 I_{\mathcal{K}_{\mathcal{Q}^*}} \subset \{v : \langle \alpha_0, v \rangle < 0\}$. Since $\mathcal{L}_0 I_{\mathcal{K}_{\mathcal{Q}^*}}$ is a closed convex cone, it follows that $\{v : \langle \alpha_0, v \rangle < 0\} \cup \{0\}$ is itself a δ -cone about $\mathcal{L}_0 I_{\mathcal{K}_{\mathcal{Q}^*}}$ which is disjoint from A , by Lemma 3.2(i). Thus this hypothesis is satisfied.

We have arrived at the following.

THEOREM 3.3. *Let A be a convex open set intersecting $\text{int } \mathfrak{S}$. Assume that $\text{dom } \Lambda_{\mathcal{P}}$ is open and (H1)–(H3) and (\mathfrak{M}') are satisfied. Let x_0 denote the initial state of $\{X_n\}$; and let \mathcal{Q}^* be the kernel defined in (3.18). Suppose that simulation is performed using a kernel $\mathcal{Q} \in \mathcal{C}_0$. Then the following hold:*

(i) *For φ a.e. x_0 ,*

$$(3.23) \quad \liminf_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2) \geq \lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}(\mathcal{E}_{\mathcal{Q}^*, \varepsilon}^2).$$

Moreover, if there is equality in (3.23), then $\mathcal{Q} = \mathcal{Q}^$ for \mathcal{P} a.e. (y, s) and φ a.e. x . Conversely, if $\mathcal{Q} = \mathcal{Q}^*$ for \mathcal{P} a.e. (y, s) , all $x \in \mathfrak{S}$, then there is equality in (3.23). Thus \mathcal{Q}^* is essentially the unique kernel in \mathcal{C}_0 which minimizes $\mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2)$ as $\varepsilon \rightarrow 0$.*

(ii) *For φ a.e. x_0 ,*

$$(3.24) \quad \lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}(\mathcal{E}_{\mathcal{Q}^*, \varepsilon}^2) = -2 \inf_{v \in A} I_{\mathcal{P}}(v).$$

Equations (3.24) and (1.4) imply that simulation performed under the distribution \mathcal{Q}^* has “logarithmic efficiency” [Asmussen (1999), page 46]. Moreover, by Remark 3.2 and a sharp form of (1.4)—available for the case that $\{\xi_n\}$ is i.i.d. and A is convex with smooth boundary [Borovkov (1997)]—one actually obtains the stronger property of “bounded relative error,” that is,

$$\limsup_{\varepsilon \rightarrow 0} \frac{\text{Var}(\mathcal{E}_{\mathcal{Q}, \varepsilon})}{\mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon})^2} < \infty.$$

It is natural to expect that this stronger property also holds at least under (\mathfrak{R}) . This property of “bounded relative error” is the strongest known property for nontrivial rare event simulation problems; see Asmussen (1999).

In the next theorem, we show [under (\mathfrak{R})] that the optimality of $\mathcal{Q}^* \in \mathcal{C}_0$ is, in fact, more general and extends to time-dependent simulation regimes of a larger class $\mathcal{C}_{\mathcal{Q}}$, defined as follows.

DEFINITIONS. (i) We say that a family of probability measures $\{\mathcal{Q}^{n,\varepsilon}(x, E \times \Gamma) : E \in \mathcal{E}, \Gamma \in \mathcal{R}^d\}$ belongs to the class \mathcal{C} if $\mathcal{Q}^{n,\varepsilon}(x, \cdot) = \mathcal{Q}^{(n\varepsilon)}(x, \cdot)$ for all $n \geq 0$, $\varepsilon > 0$, $x \in \mathbb{S}$, for some family $\{\mathcal{Q}^{(t)}(x, E \times \Gamma) : E \in \mathcal{E}, \Gamma \in \mathcal{R}^d\}$; and

$$\mathcal{P}(x, \cdot) \ll \mathcal{Q}^{(t)}(x, \cdot) \quad \text{for all } x \in \mathbb{S} \text{ and } t \geq 0.$$

(ii) We say that a family $\{\mathcal{Q}^{n,\varepsilon}(x, E \times \Gamma) : E \in \mathcal{E}, \Gamma \in \mathcal{R}^d\}$ belongs to the class \mathcal{C}_ϱ if it belongs to \mathcal{C} , and

$$\mathcal{Q}^{(t)}(x, \cdot) = \mathcal{Q}^{(\varrho - \Delta)}(x, \cdot) \quad \text{for all } x \in \mathbb{S} \text{ and } t \geq \varrho - \Delta,$$

where ϱ is the constant given in Lemma 3.2(v) and Δ is any positive constant.

The significance of the constant ϱ is made clear in Theorem 2 of Collamore (1998), where it is shown that asymptotically $\varepsilon T^\varepsilon(A) \rightarrow \varrho$ in probability, conditioned on $\{T^\varepsilon(A) < \infty\}$; that is, ϱ/ε is the “most likely” first passage time of the process $\{(X_n, S_n) : n = 0, 1, \dots\}$ into the set A/ε .

We note that the scaling of the form $\mathcal{Q}^{n,\varepsilon} = \mathcal{Q}^{(n\varepsilon)}$ coincides with the standard large deviations scaling appearing in Donsker and Varadhan (1975, 1976, 1983), Freidlin and Wentzell (1984) and essentially all subsequent work; it appeared in the context of the present problem in Collamore (1998).

For notational convenience, we will from now on write “ $\mathcal{Q} \in \mathcal{C}$ ” to mean that the family $\{\mathcal{Q}^{n,\varepsilon}(x, E \times \Gamma) : E \in \mathcal{E}, \Gamma \in \mathcal{R}^d\}$ belongs to the class \mathcal{C} , and likewise for members of \mathcal{C}_ϱ .

DEFINITION. Let $\mathcal{Q} \in \mathcal{C}$. Then we say that t_0 is a *continuity point* of \mathcal{Q} if for any $\Delta > 0$ there exists a positive constant γ such that, for all $|t - t_0| \leq \gamma$,

$$\mathcal{Q}^{(t)}(x, \cdot) \ll \mathcal{Q}^{(t_0)}(x, \cdot) \quad \text{for all } x \in \mathbb{S},$$

and for any n and ε with $|n\varepsilon - t_0| \leq \gamma$,

$$(3.25) \quad \mathbf{E}_{\pi^*} \left[\log \left(\frac{d\mathcal{Q}^{(t)}}{d\mathcal{Q}^{(t_0)}}(X_n^*, X_{n+1}^* \times \xi_{n+1}^*) \right) \right] \leq \Delta,$$

$$(3.26) \quad \mathbf{E}_x \left[\log \left(\frac{d\mathcal{Q}^{(t)}}{d\mathcal{Q}^{(t_0)}}(X_n^*, X_{n+1}^* \times \xi_{n+1}^*) \right) \right] \leq K < \infty \quad \text{for all } x \in \mathbb{S},$$

where $\{(X_n^*, S_n^*) : n = 0, 1, \dots\}$ denotes a Markov additive process having the transition kernel \mathcal{Q}^* in (3.18) and π^* is the stationary measure of $\{X_n^*\}$.

For example, if (\mathfrak{R}) holds and $\mathcal{Q} \in \mathcal{C}$ has the form

$$(3.27) \quad \mathcal{Q}^{(t)}(x, dy \times ds) = e^{(\alpha_t, s) - \Lambda_{\mathcal{P}}(\alpha_t)} \frac{r_{\mathcal{P}}(y; \alpha_t)}{r_{\mathcal{P}}(x; \alpha_t)} \mathcal{P}(x, dy \times ds),$$

where $\alpha_t = f(t)$ for some continuous function $f: [0, \infty) \rightarrow \text{int}(\text{dom } \Lambda_{\mathcal{P}})$, then all points $t \in [0, \infty)$ are continuity points of \mathcal{Q} .

DEFINITION. We say that $A \subset \mathbb{R}^d$ is a *semicone* if $v \in \partial A \Rightarrow \{\zeta v : \zeta > 1\} \subset \text{int } A$; that is, the ray generated by any point on the relative boundary of A is an *interior ray* of A .

THEOREM 3.4. *Let A be a convex open semicone intersecting $\text{int } \mathfrak{S}$. Assume that $\text{dom } \Lambda_{\mathcal{P}}$ is open and (H2) and (\mathfrak{R}) are satisfied. Let x_0 denote the initial state of $\{X_n\}$, and let $\mathcal{Q}^* \in \mathcal{C}_0$ be the kernel defined in (3.18). If simulation is performed using a family of measures $\mathcal{Q} \in \mathcal{C}$, then, for φ a.e. x_0 ,*

$$(3.28) \quad \liminf_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2) \geq \lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}(\mathcal{E}_{\mathcal{Q}^*, \varepsilon}^2).$$

Moreover, if we do not have $\mathcal{Q}^{(t_0)} = \mathcal{Q}^*$, \mathcal{P} a.e. (y, s) , φ a.e. x , at all continuity points of \mathcal{Q} in $[0, \varrho]$, then there is strict inequality in (3.28). Thus, \mathcal{Q}^* is essentially the unique element of \mathcal{C}_ϱ which minimizes $\mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2)$ as $\varepsilon \rightarrow 0$.

If $\mathcal{Q} \neq \mathcal{Q}^*$ at a continuity point t_0 which is outside $[0, \varrho]$, then we do not necessarily obtain strict inequality in (3.28); thus, the logarithmic-level optimality of \mathcal{Q}^* in Theorem 3.4 cannot be extended from \mathcal{C}_ϱ to \mathcal{C} .

REMARK 3.4. In fact what needs to be minimized in the above discussion is the number of *random variables* that need to be generated, that is,

$$(3.29) \quad \varepsilon \log[\text{Var}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2) \mathbf{E}_{\mathcal{Q}}(T^\varepsilon(A))] \quad \text{as } \varepsilon \rightarrow 0$$

[cf. Siegmund (1976), page 676, or Collamore (1996b), Lemma 5.2]. However, if A is a semicone, then

$$(3.30) \quad \lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}_{\mathcal{Q}}(T^\varepsilon(A)) = 0$$

[Collamore (1996b), Lemma 5.3]. Thus simulation under \mathcal{Q}^* is efficient and optimal.

If A is not a semicone, then the situation is more complicated; in particular, we need not have (3.30) in this case. It may then be preferable to simulate with $\mathcal{Q} \in \mathcal{C}$, where $\mathcal{Q}^{(t)} = \mathcal{Q}^*$ for $t \in [0, \varrho]$, but $\mathcal{Q}^{(t)} \neq \mathcal{Q}^*$ for $t \geq t_0$, some $t_0 > \varrho$. By a judicious choice of \mathcal{Q} , one may often obtain both (3.30) and (3.24).

3.3. *General sets.* Finally, suppose that A is an arbitrary open subset of \mathbb{R}^d . In this case, we will show that A can be partitioned into subsets A_1, \dots, A_l , and that the techniques of Theorem 3.1 can be applied to efficiently simulate $\mathbf{P}\{T^\varepsilon(A) < \infty, \varepsilon S_{T^\varepsilon(A)} \in A_i\}$, for $i = 1, \dots, l$.

For any $\alpha \in \text{dom } \Lambda_{\mathcal{P}}$, let

$$\mathcal{Q}_\alpha(x, dy \times ds) = e^{(\alpha, s) - \Lambda_{\mathcal{P}}(\alpha)} \frac{r_{\mathcal{P}}(y; \alpha)}{r_{\mathcal{P}}(x; \alpha)} \mathcal{P}(x, dy \times ds).$$

Let $B \subset A$, and let $\mathfrak{P}_k^\varepsilon$ denote the paths which first hit A/ε at time k , as defined formally in (3.2). If simulation is performed using a kernel $\mathcal{Q} \in \mathcal{C}_0$, then

$$(3.31) \quad \begin{aligned} & \mathbf{P}\{T^\varepsilon(A) < \infty, \varepsilon S_{T^\varepsilon(A)} \in B\} \\ &= \sum_k \int_{\mathfrak{P}_k^\varepsilon} \left(\prod_{n=0}^{k-1} \frac{d\mathcal{P}}{d\mathcal{Q}}(x_n, x_{n+1} \times s_{n+1}) \right) \mathbb{1}_B(\varepsilon S_k) \\ & \quad \times \mathcal{Q}(x_0, dx_1 \times ds_1) \cdots \mathcal{Q}(x_{k-1}, dx_k \times ds_k). \end{aligned}$$

Hence

$$(3.32) \quad \mathcal{E}_{\mathcal{Q}, \varepsilon}(B) \stackrel{\text{def}}{=} \prod_{n=0}^{T^\varepsilon(A)-1} \left(\frac{d\mathcal{P}}{d\mathcal{Q}}(\tilde{X}_n, \tilde{X}_{n+1} \times \tilde{\xi}_{n+1}) \right) \mathbb{1}_B(\varepsilon S_{T^\varepsilon(A)})$$

is an unbiased estimator for $\mathbf{P}\{T^\varepsilon(A) < \infty, \varepsilon S_{T^\varepsilon(A)} \in B\}$, where $\{(\tilde{X}_n, \tilde{S}_n) : n = 0, 1, \dots\}$ denotes a Markov additive process having transition kernel \mathcal{Q} .

PROPOSITION 3.5. *Let $\Delta > 0$ and $A \subset \mathbb{R}^d$, and suppose that $\text{dom } \Lambda_{\mathcal{P}}$ is open and (H2) and the lower bound of (\mathfrak{R}) are satisfied. Let x_0 denote the initial state of $\{X_n\}$, and let $a = \inf_{v \in A} I_{\mathcal{P}}(v)$. Then the following hold:*

- (i) *For some finite subset $\{\alpha_1, \dots, \alpha_l\}$ of $\partial(\mathcal{L}_0 \Lambda_{\mathcal{P}})$, the collection $\{\mathcal{H}(\alpha_i, a - \Delta) : i = 1, \dots, l\}$ is an open cover for A .*
- (ii) *Let $A_1 = A \cap \mathcal{H}(\alpha_1, a - \Delta)$, $A_2 = (A \cap \mathcal{H}(\alpha_2, a - \Delta)) - A_1$, and so on for A_3, \dots, A_l . Then $\{A_1, \dots, A_l\}$ is a partition of A , and, for each i ,*

$$(3.33) \quad \mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2(A_i)) \leq C \exp\left\{-2\varepsilon^{-1} \left(\inf_{v \in A} I_{\mathcal{P}}(v) - \Delta\right) / \varepsilon\right\} \quad \varphi \text{ a.e. } x_0$$

for a certain positive constant C . In the event that A is a finite union of disjoint convex sets $\{A'_1, \dots, A'_k\}$, then instead we may take $A_i = A'_i$, and (3.33) holds with $\Delta = 0$.

REMARK 3.5. If (\mathfrak{M}') and condition (\mathcal{C}) in the statement of Theorem 3.1 are satisfied, then it is not necessary to assume that the lower bound of (\mathfrak{R}) holds. However, in this case we must replace (3.33) with logarithmic asymptotics, as given in (3.8).

4. Proofs. We now introduce some further notation from convex analysis.

For any convex function f , let f^* , $\text{cl } f$, $f^{0+}(\cdot)$, $0^+ f$, $\text{dom } f$ and $\partial f(\cdot)$ denote the convex conjugate of f , the closure of f , the recession function of f , the recession cone of f , the domain of f and the subgradient set of f , respectively.

For any convex set C , let

$$\delta_C(v) = \begin{cases} 0, & v \in C, \\ \infty, & v \notin C, \end{cases}$$

and let C° , 0^+C , $\text{aff}C$, $\text{ri}C$ and ∂C denote the polar of C , the recession cone of C , the affine hull of C , the relative interior of C and the relative boundary of C , respectively. [For definitions, see Rockafellar (1970).]

Also, we adopt the same terminology that was already introduced at the beginning of Sections 2 and 3.

4.1. *Proof of Theorem 3.1: upper bound.* The proof of the upper bound is based on the following convexity lemma, which shows that the separation property described in Lemma 3.2(i) is in fact quite general.

LEMMA 4.1. *Let $A \subset \mathbb{R}^d$ be a convex open set, and let f be a closed convex function. Let $I(v) = \sup\{\langle \alpha, v \rangle : \alpha \in \mathcal{L}_0 f\}$ and $a = \inf_{v \in A} I(v)$. Assume that A intersects $\text{dom} I$, but $A \cap (\mathcal{L}_0 f)^\perp = \emptyset$. Then there exists $\theta \in \mathcal{L}_0 f$ such that*

$$(4.1) \quad A \subset \{v : \langle \theta, v \rangle > a\} \quad \text{and} \quad \mathcal{L}_a I \subset \{v : \langle \theta, v \rangle \leq a\}.$$

PROOF. Note that $(\mathcal{L}_0 f)^\perp = \mathcal{L}_0 I$. Since A does not intersect this zero-set and I is a positively homogeneous convex function, the sets A and $\mathcal{L}_a I$ are convex with no common points in their relative interiors. Hence there exists a separating hyperplane [Rockafellar (1970), Theorem 11.3], that is, for some $\beta \in \mathbb{R}^d - \{0\}$,

$$(4.2) \quad A \subset \{v : \langle \beta, v \rangle > b\} \quad \text{and} \quad \mathcal{L}_a I \subset \{v : \langle \beta, v \rangle \leq b\},$$

where $b \in \mathbb{R}$; in fact, $b \geq 0$ because the definition of I implies $I(0) = 0$, so $0 \in \mathcal{L}_a I$.

Let $c > 0$, and define $J = I - c$. Then $\mathcal{L}_0 J = \mathcal{L}_c I$, and $J^* = \delta_{\mathcal{L}_0 f} + c$ [Rockafellar (1970), Theorem 12.2]. An application of Theorems 13.5 and 9.7 of Rockafellar (1970) then gives

$$(4.3) \quad \delta_{\mathcal{L}_c I}^*(\beta) = \inf\{(J^* \gamma)(\beta) \mid \gamma > 0 \text{ or } \gamma = 0^+\},$$

where $(J^* \gamma)(\cdot) = \gamma J^*(\cdot/\gamma)$, for all $\gamma > 0$, and $J^* 0^+$ is the recession function of J^* .

Note that $J^*(\cdot) \in \{c, \infty\} \Rightarrow J^* 0^+(\cdot) \in \{0, \infty\}$ [Rockafellar (1970), Theorem 8.5]. However, we cannot have $J^* 0^+(\beta) = 0$. Otherwise, (4.3) would imply $\delta_{\mathcal{L}_c I}^*(\beta) = 0$ for all $c > 0$; then $\mathcal{L}_c I \subset \{v : \langle \beta, v \rangle \leq 0\}$ for all $c > 0$; and since A is open, we would then obtain $A \cap \text{dom} I = \emptyset$, by (4.2), which is contrary to hypothesis. Therefore $J^* 0^+(\beta) = \infty$; thus the point $\gamma = 0^+$ can be removed from the infimum in (4.3).

Since $J^*(\cdot) \in \{c, \infty\}$, we now conclude from (4.3) that, for all $c > 0$,

$$(4.4) \quad \delta_{\mathcal{L}_c I}^*(\beta) = \gamma c, \quad \text{where } \gamma = \inf \left\{ \tilde{\gamma} : \frac{\beta}{\tilde{\gamma}} \in \mathcal{L}_0 f \right\}.$$

Setting $c = a$ when $a > 0$ yields $\mathcal{L}_a I \subset \{v : \langle \beta, v \rangle \leq \gamma a\}$. Since $b \geq 0$, we conclude that the constant b in (4.2) is greater than or equal to γa , for any $a \geq 0$.

Next suppose $b \geq \gamma a'$, where $a' > a$. Then $\mathcal{L}_{a'} I \subset \{v : \langle \beta, v \rangle \leq b\}$. By (4.2) it follows that $\inf_{v \in A} I(v) \geq a'$, which contradicts the definition of a . Therefore $b = \gamma a$. Also observe that $\gamma > 0$, because otherwise (4.2) and (4.4) would once again imply $A \cap \text{dom } I = \emptyset$, contrary to hypothesis. The required result now follows from (4.2) by setting $\theta = \beta/\gamma$. \square

Proof of Theorem 3.1 (Upper bound). If $\mathcal{Q} = \mathcal{P}$, then the result follows trivially from (1.4); we will assume from now on that this is not the case.

First assume that condition (\mathfrak{C}) in the statement of the theorem is satisfied. Under this general assumption, the eigenvectors $r_{\mathcal{K}_{\mathcal{Q}}}$ need not be uniformly positive. Consequently we start by introducing an augmented kernel, $\mathcal{K}_{\mathcal{Q}}^{\Delta}$, whose eigenvectors *do* have this positivity property. Namely, for any $\Delta > 0$ define

$$(4.5) \quad \mathcal{K}_{\mathcal{Q}}^{\Delta}(x, dy \times ds) = \mathcal{K}_{\mathcal{Q}}(x, dy \times ds) + \Delta \eta_{x_0}(dy) \eta_{\emptyset}(ds),$$

where η_{x_0} denotes a measure on \mathbb{S} having point mass at x_0 and η_{\emptyset} denotes a measure on \mathbb{R}^d having point mass at the origin. For shorthand notation, let $(\lambda_{\Delta}(\alpha))^{-1}$ denote the convergence parameter of $\mathcal{K}_{\mathcal{Q}}^{\Delta}(\alpha)$ and $\Lambda_{\Delta}(\alpha) = \log \lambda_{\Delta}(\alpha)$.

We begin by establishing the following.

ASSERTION. $\Gamma_0(\alpha) \stackrel{\text{def}}{=} \lim_{\Delta \rightarrow 0} \Lambda_{\Delta}(\alpha) \leq \bar{\Lambda}_{\mathcal{K}_{\mathcal{Q}}}(\alpha)$, for all $\alpha \in \mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_{\mathcal{Q}}}$.

PROOF. Note

$$(4.6) \quad \begin{aligned} & (\hat{\mathcal{K}}_{\mathcal{Q}}^{\Delta})^k(x_0, \mathbb{S}; \alpha) \\ &= \sum_{(i_1, \dots, i_k) \in \mathfrak{J}} \int_{\mathbb{S}^k \times \mathbb{R}^d} e^{\langle \alpha, s \rangle} \mathfrak{J}^{(i_1)}(x_0, dx_1 \times \cdot) * \dots * \mathfrak{J}^{(i_k)}(dx_{k-1}, dx_k \times \cdot), \end{aligned}$$

where $\mathfrak{J}^{(0)} = \mathcal{K}_{\mathcal{Q}}$ and $\mathfrak{J}^{(1)} = \Delta \eta_{x_0} \eta_{\emptyset}$, and \mathfrak{J} consists of all elements of the form (i_1, \dots, i_k) , where either $i_j = 0$ or 1. Fix $\alpha \in \mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_{\mathcal{Q}}}$ and $N \in \mathbb{Z}_+$, and let

$$\log b_N = N[\Lambda_{\mathcal{K}_{\mathcal{Q}}}^{(1)}(\alpha) - \Lambda_{\mathcal{K}_{\mathcal{Q}}}^{(N)}(\alpha)] \quad [< \infty \text{ by (H1)}].$$

Observe that any product $\mathfrak{J}^{(1)} * \dots * \mathfrak{J}^{(k)}$ which has n “ $\Delta \eta_{x_0} \eta_{\emptyset}$ ” terms contains at most $n + 1$ products consisting of j consecutive “ $\mathcal{K}_{\mathcal{Q}}^{(i)}$ ” terms where $j < N$. Consequently, for a product containing n “ $\Delta \eta_{x_0} \eta_{\emptyset}$ ” terms,

$$(4.7) \quad \begin{aligned} & \int_{\mathbb{S}^k \times \mathbb{R}^d} e^{\langle \alpha, s \rangle} \mathfrak{J}^{(i_1)}(x_0, dx_1 \times \cdot) * \dots * \mathfrak{J}^{(i_k)}(dx_{k-1}, dx_k \times \cdot) \\ & \leq b_N^{n+1} \Delta^n (\lambda_{\mathcal{K}_{\mathcal{Q}}}^{(N)}(\alpha))^{k-n}. \end{aligned}$$

Summing all terms in (4.6) gives $\lambda_{\Delta}(\alpha) \leq \lambda_{\mathcal{K}_{\mathcal{Q}}}^{(N)}(\alpha) + b_N \Delta$, by the definition of the convergence parameter. Letting $\Delta \rightarrow 0$ and then $N \rightarrow \infty$ establishes the assertion. \square

Let $\Gamma = \text{cl}\Gamma_0$. We now apply Lemma 4.1 with $f(\alpha) = \Gamma(\alpha) - c$, $c < 0$, and $I = J^{(c)}$, where

$$J^{(c)}(v) = \sup\{\langle \alpha, v \rangle : \alpha \in \mathcal{L}_c\Gamma\} \quad \text{for any } c \in \mathbb{R}.$$

Assume, for the moment, that the assumptions of the lemma are satisfied (we will verify these later), and let θ be the element obtained in the lemma when $f(\alpha) = \Gamma(\alpha) - c$. Then $\Gamma(\theta) \leq c$; and by the assertion, there exists a $\Delta > 0$ such that $\Lambda_\Delta(\theta) \leq 0$. Moreover, there exists a $(\lambda_\Delta(\theta))^{-1}$ -subinvariant function $r_\Delta(\theta)$ for the kernel $\hat{\mathcal{K}}_\mathcal{Q}^\Delta(\theta)$ [Nummelin (1984), Proposition 5.2 and Theorem 5.1, or Section 2.2 above]. Define

$$(4.8) \quad \mathcal{R}_\theta(x, dy \times ds) = \frac{e^{\langle \theta, s \rangle} r_\Delta(dy; \theta)}{r_\Delta(x; \theta)} \mathcal{K}_\mathcal{Q}^\Delta(x, dy \times ds).$$

Since $r_\Delta(\theta)$ is $(\lambda_\Delta(\theta))^{-1}$ -subinvariant and $\Lambda_\Delta(\theta) \leq 0$, \mathcal{R}_θ is itself a Markov additive subprobability kernel.

Let $\mathfrak{P}_k^\varepsilon$ denote the paths which first hit A/ε at time $T^\varepsilon(A) = k$. Then by (3.4) and the definition of $\mathcal{K}_\mathcal{Q}^\Delta$,

$$(4.9) \quad \begin{aligned} \mathbf{E}(\mathfrak{E}_{\mathcal{Q}, \varepsilon}^2) &\leq \sum_{k=1}^\infty \int_{\mathfrak{P}_k^\varepsilon} \mathcal{K}_\mathcal{Q}^\Delta(x_0, dx_1 \times ds_1) \cdots \mathcal{K}_\mathcal{Q}^\Delta(x_{k-1}, dx_k \times ds_k) \\ &= \sum_{k=1}^\infty \int_{\mathfrak{P}_k^\varepsilon} \frac{r_\Delta(x_0; \theta) e^{-\langle \theta, s_1 + \cdots + s_k \rangle}}{r_\Delta(x_k; \theta)} \\ &\quad \times \mathcal{R}_\theta(x_0, dx_1 \times ds_1) \cdots \mathcal{R}_\theta(x_{k-1}, dx_k \times ds_k). \end{aligned}$$

Note that $r_\Delta(\cdot; \theta)$ is uniformly positive [Remark 2.3, since (4.5) yields a minorization $\Delta\eta_{x_0}\eta_\theta$]. Also, $r_\Delta(\cdot; \theta) < \infty$ φ a.e. [Nummelin (1984), Proposition 5.1]. Thus the ratio $(r_\Delta(x_0; \theta)/r_\Delta(x_k; \theta))$ in (4.9) is deterministically bounded.

Next observe by Lemma 4.1 that $\langle \theta, S_{T^\varepsilon(A)} \rangle > \inf\{J^{(c)}(v) : v \in A\}/\varepsilon$. Hence the integrand in (4.9) is less than or equal to $\text{const} \cdot \exp\{-\inf_{v \in A} J^{(c)}(v)/\varepsilon\}$. Since \mathcal{R}_θ is a subprobability kernel, we then obtain upon letting $\varepsilon \searrow 0$ in (4.9) that

$$(4.10) \quad \limsup_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}(\mathfrak{E}_{\mathcal{Q}, \varepsilon}^2) \leq - \inf_{v \in A} J^{(c)}(v).$$

It remains to show

$$(4.11) \quad \lim_{c \rightarrow 0} \left\{ \inf_{v \in A} J^{(c)}(v) \right\} \geq \inf_{v \in \mathfrak{A}} \bar{I}_{\mathcal{K}_\mathcal{Q}}(v).$$

Observe that $\Gamma(0) > 0$ [Lemma 2.3(ii), since $\mathcal{P} \neq \mathcal{Q}$]. Hence $(\mathcal{L}_c\Gamma)^\perp \searrow (\mathcal{L}_0\Gamma)^\perp$ as $c \nearrow 0$. It follows under (C) that

$$(4.12) \quad A \cap \text{cone}_{\delta'}(\mathcal{L}_c\Gamma)^\perp = \emptyset \quad \text{for all } c \in [c_0, 0), \text{ for some } \delta' > 0 \text{ and } c_0 < 0.$$

Since $(\mathcal{L}_c \Gamma)^\perp = \mathcal{L}_0 J^{(c)}$, this implies

$$(4.13) \quad A \cap \text{cone}_{\delta'}(\mathcal{L}_0 J^{(c)}) = \emptyset \quad \text{for all } c \in [c_0, 0).$$

Now $J^{(c)}$ is bounded away from zero on $(\text{cone}_{\delta'}(\mathcal{L}_0 J^{(c)}))^c \cap \mathbf{S}^{d-1}$. Since $J^{(c)}$ is positively homogeneous, it follows that its level sets are compact on the restricted set $(\text{cone}_{\delta'}(\mathcal{L}_0 J^{(c)}))^c$. Hence, if a denotes the limit on the left-hand side of (4.11), then the sets $\text{cl } A \cap \mathcal{L}_a J^{(c)}$ are compact and decrease monotonically as $c \nearrow 0$ to

$$\text{cl } A \cap \mathcal{L}_a J^{(0)} \subset \text{cl } A \cap \mathcal{L}_a \bar{I}_{\mathcal{K}_Q}.$$

Since the left-hand side is nonempty, we conclude

$$(4.14) \quad \lim_{c \rightarrow 0} \left\{ \inf_{v \in A} J^{(c)}(v) \right\} \geq \inf_{v \in \text{cl } A} \bar{I}_{\mathcal{K}_Q}(v).$$

Also, since $\bar{I}_{\mathcal{K}_Q}$ is positively homogeneous,

$$(4.15) \quad \inf\{\bar{I}_{\mathcal{K}_Q}(v) : v \in \text{cl } A\} = \inf\{\bar{I}_{\mathcal{K}_Q}(v) : v \in \mathfrak{A}\},$$

provided that the infimum on the left is greater than or equal to 0. As $\mathcal{L}_0 \bar{I}_{\mathcal{K}_Q} = (\mathcal{L}_0 \bar{\Lambda}_{\mathcal{K}_Q})^\perp$ is disjoint from A , we see that this infimum is indeed nonnegative. Consequently, (4.11) follows from (4.14) and (4.15).

Finally, we need to verify that the conditions of Lemma 4.1 are actually satisfied when $f(\alpha) = \Gamma(\alpha) - c$, for $c < 0$ sufficiently large. To this end, note that Γ is convex, since [by Lemma 4.2(ii) below] it is the limit of convex functions. Also, if $c < 0$ is sufficiently large, then by Lemma 4.3(i) below,

$$\text{ri } \mathfrak{D}_{\mathcal{K}_Q}^{(c)} \supset \text{ri } \mathfrak{D}_{\mathcal{P}} = \text{ri } \mathfrak{S}.$$

[The last step follows from Theorem 13.5 of Rockafellar (1970), as noted in the remark following Theorem 2.1 of Collamore (1996a).] Then $\text{dom } J^{(c)} \supset \mathfrak{D}_{\mathcal{K}_Q}^{(c)} \Rightarrow A \cap \text{dom } J^{(c)} \neq \emptyset$. Since (4.12) holds, we conclude that the hypotheses of the lemma are satisfied for sufficiently large $c < 0$.

If the lower bound of (\mathfrak{R}) is satisfied [instead of condition (\mathfrak{C})], then we may apply the measure transformation (4.8) *directly* with \mathcal{K}_Q in place of \mathcal{K}_Q^Δ , and Lemma 4.1 directly with $f = \Lambda_{\mathcal{K}_Q}$ and $I = I_{\mathcal{K}_Q}$. The uniform positivity of $r_{\mathcal{K}_Q}$ is obtained from Remark 2.3.

4.2. Proof of Theorem 3.1: lower bound. We begin by introducing a splitting and truncation of \mathcal{K}_Q , as follows.

Let $h \otimes \nu$ be a minorization as in (\mathfrak{M}') , with $h < 1$, and note under (\mathfrak{M}') that either $\nu(dy \times ds) = \nu(dy)\eta_\emptyset(ds)$ or $h(x, ds) = h(x)\eta_\emptyset(ds)$, where η_\emptyset is a point mass at the origin. Thus $(h \otimes \nu(\cdot))^2 = g \otimes \mu(\cdot)$, where $g = h^2$ and $\mu = \nu^2$. Hence by Lemma 2.3(i), $(g \otimes \nu) \leq \mathcal{K}_Q$. This implies the minorization

$$(4.16) \quad \hat{g}(\alpha) \otimes \hat{\mu}(\alpha) \leq \hat{\mathcal{K}}_Q(\alpha) \quad \text{for all } \alpha.$$

Define

$$\bar{\mathcal{K}}_{\mathcal{Q}}(x, dy \times ds) = \frac{\mathcal{K}_{\mathcal{Q}}(x, dy \times ds) - (g \otimes \mu)(x, dy \times ds)}{1 - g(x, \mathbb{R}^d)} \quad (\geq 0),$$

and observe by this definition that

$$(4.17) \quad \mathcal{K}_{\mathcal{Q}}(x, dy \times ds) = (g \otimes \mu)(x, dy \times ds) + (1 - g(x, \mathbb{R}^d))\bar{\mathcal{K}}_{\mathcal{Q}}(x, dy \times ds).$$

Enlarge $(\mathbb{S}, \mathfrak{f})$ to $(\tilde{\mathbb{S}}, \tilde{\mathfrak{f}})$, where $\tilde{\mathbb{S}} = \mathbb{S} \times \{0, 1, 2, \dots\}$ and $\tilde{\mathfrak{f}}$ is the natural extension of \mathfrak{f} to $\tilde{\mathbb{S}}$; and for $M \in \mathbb{Z}_+$ define truncated versions $g_M, h_M, \bar{\mathcal{K}}_{\mathcal{Q}}^M, \mathcal{K}_{\mathcal{Q}}^M$ by

$$\begin{aligned} g_M((x, i), ds) &= \left(\frac{M}{M+1} \right) \mathbb{1}_0(i) \mathbb{1}_{[-M, M]^d}(s) g(x, ds), \\ \mu_M((dy, j), ds) &= \mathbb{1}_1(j) \mathbb{1}_{[-M, M]^d}(s) \mu(dy \times ds), \\ \bar{\mathcal{K}}_{\mathcal{Q}}^M((x, i), (dy, j) \times ds) &= \frac{M}{M+1} [\mathbb{1}_{i+1}(j) \mathbb{1}_{(0, M]}(j) \mathbb{1}_{[-M, M]^d}(s) \\ &\quad \times \bar{\mathcal{K}}_{\mathcal{Q}}(x, dy \times ds)] \wedge M, \\ \mathcal{K}_{\mathcal{Q}}^M((x, i), (dy, j) \times ds) &= g_M \otimes \mu_M((x, i), (dy, j) \times ds) \\ &\quad + (1 - g(x, \mathbb{R}^d)) \bar{\mathcal{K}}_{\mathcal{Q}}^M((x, i), (dy, j) \times ds). \end{aligned}$$

Note that $\mathcal{K}_{\mathcal{Q}}^M$ is strictly increasing,

$$(4.18) \quad \mathcal{K}_{\mathcal{Q}}^M((x \times \mathbb{N}), (E \times \mathbb{N}) \times \Gamma) \leq \mathcal{K}_{\mathcal{Q}}(x, E \times \Gamma) \quad \text{for all } E \in \mathfrak{f}, \Gamma \in \mathcal{R}^d$$

(where \mathbb{N} denotes the set of natural numbers), and

$$(4.19) \quad \mathcal{K}_{\mathcal{Q}}^M((x \times \mathbb{N}), (E \times \mathbb{N}) \times \Gamma) \nearrow \mathcal{K}_{\mathcal{Q}}(x, E \times \Gamma) \quad \text{as } M \rightarrow \infty.$$

Also, it follows from our construction that $\mathcal{K}_{\mathcal{Q}}^M$ is irreducible with respect to a maximal irreducibility measure $\varphi_M \nearrow \varphi$ as $M \rightarrow \infty$.

The kernel $\mathcal{K}_{\mathcal{Q}}^M$ has a minorization, namely $g_M \otimes \mu_M \leq \mathcal{K}_{\mathcal{Q}}^M$, which implies

$$(4.20) \quad \hat{g}_M(\alpha) \otimes \hat{\mu}_M(\alpha) \leq \hat{\mathcal{K}}_{\mathcal{Q}}^M(\alpha) \quad \text{for all } \alpha.$$

For shorthand notation, let $(\lambda_M(\alpha))^{-1}$ denote the convergence parameter of $\hat{\mathcal{K}}_{\mathcal{Q}}^M(\alpha)$, $\Lambda_M(\alpha) = \log \lambda_M(\alpha)$, $\mathfrak{D}_M = \mathfrak{D}_{\mathcal{K}_{\mathcal{Q}}^M}$ and $I_M^{(c)} = I_{\mathcal{K}_{\mathcal{Q}}^M}^{(c)}$.

Our main reason for introducing the above truncation is because the transformed kernel $\hat{\mathcal{K}}_{\mathcal{Q}}^M(\alpha)$ has eigenvectors which are *bounded*, and the following regularity properties also hold.

LEMMA 4.2. *Let $\mathcal{K}_{\mathcal{Q}}^M, \mathcal{K}_{\mathcal{Q}}, \lambda_M, \Lambda_M$ and $\Lambda_{\mathcal{K}_{\mathcal{Q}}}$ be defined as above. Then the following hold:*

- (i) Λ_M is convex, analytic, strictly increasing, and $\Lambda_M(\alpha) \nearrow \Lambda_{\mathcal{K}_{\mathcal{Q}}}(\alpha)$ as $M \rightarrow \infty$, for all α .
- (ii) $\Lambda_{\mathcal{K}_{\mathcal{Q}}}$ is convex and lower semicontinuous.

(iii) For any α , there exists a $(\lambda_M(\alpha))^{-1}$ -invariant function, $r_M(\cdot; \alpha)$, for the kernel $\hat{\mathcal{K}}_{\mathcal{Q}}^M(\alpha)$. Moreover, the function $r_M(\cdot; \alpha)$ is positive and bounded.

PROOF. (i) Following Iscoe, Ney and Nummelin [(1985), Lemma 3.4], introduce the generating function

$$\begin{aligned} &\psi_M(\alpha, \zeta) \\ (4.21) \quad &= \sum_{n=1}^{\infty} \int_{s \in \mathbb{R}^d, x, y \in \tilde{\mathbb{S}}} e^{(\alpha, s) - \zeta n} (\mu_M \otimes (\mathcal{K}_{\mathcal{Q}}^M - g_M \otimes \mu_M)^{n-1} \otimes g_M)(dx, dy \times ds) \\ &= \sum_{n=1}^{\infty} e^{-\zeta n} \hat{\mu}_M(\alpha) (\hat{\mathcal{K}}_{\mathcal{Q}}^M(\alpha) - \hat{g}_M(\alpha) \otimes \hat{\mu}_M(\alpha))^{n-1} \hat{g}_M(\alpha). \end{aligned}$$

Then $\Lambda_M(\alpha) = \inf\{\zeta : \psi_M(\alpha, \zeta) < 1\}$ [Nummelin (1984), Proposition 4.7(i)]. Note by the construction of $\mathcal{K}_{\mathcal{Q}}^M$ that the individual terms and number of nonzero terms in the summand on the right-hand side of (4.21) are finite; consequently,

$$(4.22) \quad \psi_M(\alpha, \Lambda_M(\alpha)) = 1.$$

The convexity of Λ_M follows from (4.22) and the convexity of ψ_M . Since ψ_M is analytic on \mathbb{R}^{d+1} , the analyticity of Λ_M follows from (4.22) and the implicit function theorem. Finally, the convergence $\Lambda_M \nearrow \Lambda_{\mathcal{K}_{\mathcal{Q}}}$ is obtained as in Ney and Nummelin [(1987b), Lemma 3.3(i)]. (From this argument, we also see that Λ_M is strictly increasing, since $\mathcal{K}_{\mathcal{Q}}^M$ is.)

(ii) $\Lambda_{\mathcal{K}_{\mathcal{Q}}}$ is convex because [by (i)] it is a limit of convex functions, and lower semicontinuous since the analytic functions $\Lambda_M \nearrow \Lambda_{\mathcal{K}_{\mathcal{Q}}}$ as $M \rightarrow \infty$.

(iii) Since (4.22) holds, $\hat{\mathcal{K}}_{\mathcal{Q}}^M$ is $(\lambda_M(\alpha))^{-1}$ -recurrent [Nummelin (1984), Proposition 4.3]. Hence a $(\lambda_M(\alpha))^{-1}$ -invariant function exists and is given by

$$(4.23) \quad r_M(\alpha) = \sum_{n=0}^{\infty} e^{-(n+1)\Lambda_M(\alpha)} (\hat{\mathcal{K}}_{\mathcal{Q}}^M(\alpha) - \hat{g}_M(\alpha) \otimes \hat{\mu}_M(\alpha))^n \hat{g}_M(\alpha)$$

[Nummelin (1984), Theorem 5.1]. By the construction of $\mathcal{K}_{\mathcal{Q}}^M$, the sum and individual terms on the right-hand side are finite; hence $r_M(\cdot; \alpha)$ is bounded. The positivity of $r_M(\cdot; \alpha)$ is obtained from Nummelin [(1984), Proposition 5.1]. \square

LEMMA 4.3. Let the kernels \mathcal{P} , $\mathcal{K}_{\mathcal{Q}}^M$ and $\mathcal{K}_{\mathcal{Q}}$ be defined as above. Then the following hold:

- (i) If $\mathcal{L}_b \Lambda_{\mathcal{P}} \neq \emptyset$ and $\mathcal{L}_c \Lambda_{\mathcal{K}_{\mathcal{Q}}} \neq \emptyset$, then $\mathcal{D}_{\mathcal{P}}^{(b)} \subset \mathcal{D}_{\mathcal{K}_{\mathcal{Q}}}^{(c)}$.
- (ii) If $\mathcal{L}_{c_M} \Lambda_M \neq \emptyset$ and $\mathcal{L}_c \Lambda_{\mathcal{K}_{\mathcal{Q}}} \neq \emptyset$, then $\mathcal{D}_M^{(c_M)} \nearrow \mathcal{D}_{\mathcal{K}_{\mathcal{Q}}}^{(c)}$ as $M \rightarrow \infty$.
- (iii) If $\mathcal{L}_c \Lambda_{\mathcal{K}_{\mathcal{Q}}} \neq \emptyset$ and $v \in \text{ri } \mathcal{D}_{\mathcal{K}_{\mathcal{Q}}}^{(c)}$, then $I_M^{(c)}(v) \searrow I_{\mathcal{K}_{\mathcal{Q}}}^{(c)}(v)$ as $M \rightarrow \infty$.
- (iv) If $\mathcal{L}_c \Lambda_M \neq \emptyset$ and $v \in \text{ri } \mathcal{D}_M^{(c)}$, then the supremum in the definition of $I_M^{(c)}$ is achieved at a point $\theta \in \partial(\mathcal{L}_c \Lambda_M) \cap \text{aff } \mathcal{D}_M^{(c)}$. Moreover, if $c > \inf_{\alpha} \Lambda_M(\alpha)$, then $\rho \nabla \Lambda_M(\theta) = v$ for some nonnegative constant ρ .

PROOF. (i) Note

$$(4.24) \quad (\mathfrak{D}_{\mathcal{P}}^{(b)})^\circ = 0^+(\mathcal{L}_b \Lambda_{\mathcal{P}}) \quad \text{and} \quad (\mathfrak{D}_{\mathcal{K}_Q}^{(c)})^\circ = 0^+(\mathcal{L}_c \Lambda_{\mathcal{K}_Q})$$

[Rockafellar (1970), Theorem 14.2, applied to $\delta_{\mathcal{L}_b \Lambda_{\mathcal{P}}}^*$ and $\delta_{\mathcal{L}_c \Lambda_{\mathcal{K}_Q}}^*$; by Rockafellar (1970), Theorem 8.7, $0^+(\delta_{\mathcal{L}_b \Lambda_{\mathcal{P}}})$, $0^+(\delta_{\mathcal{L}_c \Lambda_{\mathcal{K}_Q}})$ may be identified with $0^+(\mathcal{L}_b \Lambda_{\mathcal{P}})$, $0^+(\mathcal{L}_c \Lambda_{\mathcal{K}_Q})$, respectively]. Now set $b = c/2$. Since $\mathcal{L}_c \Lambda_{\mathcal{K}_Q} \subset 2(\mathcal{L}_{c/2} \Lambda_{\mathcal{P}})$ [Lemma 2.3(ii)], it follows from (4.24) that $(\mathfrak{D}_{\mathcal{P}}^{(c/2)})^\circ \supset (\mathfrak{D}_{\mathcal{K}_Q}^{(c)})^\circ$; hence $\mathfrak{D}_{\mathcal{P}}^{(c/2)} \subset \mathfrak{D}_{\mathcal{K}_Q}^{(c)}$. For a general $b \geq \inf_{\alpha} \Lambda_{\mathcal{P}}(\alpha)$, observe that $0^+(\mathcal{L}_b \Lambda_{\mathcal{P}}) = 0^+(\mathcal{L}_{c/2} \Lambda_{\mathcal{P}})$ [Rockafellar (1970), Theorem 8.7]; hence $\mathfrak{D}_{\mathcal{P}}^{(b)} = \mathfrak{D}_{\mathcal{P}}^{(c/2)}$.

(ii) The proof is analogous to (i), once it is observed that

$$(4.25) \quad (0^+ \Lambda_{\mathcal{K}_Q}) = \bigcap_M (0^+ \Lambda_M)$$

[Rockafellar (1970), Corollary 8.3.3 and Theorem 8.7].

(iii) First assume $v \in \text{int } \mathfrak{D}_{\mathcal{K}_Q}^{(c)}$. Let

$$\mathcal{W}_M = \{\alpha \in \mathcal{L}_c \Lambda_M : \langle \alpha, v \rangle \geq I_{\mathcal{K}_Q}^{(c)}(v)\}, \quad \mathcal{W} = \{\alpha \in \mathcal{L}_c \Lambda_{\mathcal{K}_Q} : \langle \alpha, v \rangle \geq I_{\mathcal{K}_Q}^{(c)}(v)\}.$$

Since $\mathcal{L}_c \Lambda_M \searrow \mathcal{L}_c \Lambda_{\mathcal{K}_Q}$ monotonically as $M \rightarrow \infty$, $\bigcap_M \mathcal{W}_M = \mathcal{W} = \partial I_{\mathcal{K}_Q}^{(c)}(v)$ [the last step follows from Theorem 23.5 of Rockafellar (1970)]. Now $v \in \text{int } \mathfrak{D}_{\mathcal{K}_Q}^{(c)} \Rightarrow \partial I_{\mathcal{K}_Q}^{(c)}(v)$ is a nonempty compact set [Rockafellar (1970), Theorem 23.4]. Hence the convergence $\mathcal{W}_M \searrow \mathcal{W}$ implies

$$\mathcal{W}_M \subset \{z : \|z - w\| < \Delta, w \in \mathcal{W}\}, \quad M \geq \text{some } M_0(\Delta), \text{ for any } \Delta > 0.$$

Thus $I_M^{(c)}(v) \leq I_{\mathcal{K}_Q}^{(c)}(v) + \Delta \|v\|$, all $M \geq M_0(\Delta)$. Conversely, $\mathcal{L}_0 \Lambda_M \supset \mathcal{L}_0 \Lambda_{\mathcal{K}_Q} \Rightarrow I_M^{(c)}(v) \geq I_{\mathcal{K}_Q}^{(c)}(v)$, for all M . We conclude that $I_M^{(c)}(v) \searrow I_{\mathcal{K}_Q}^{(c)}(v)$.

Next assume $v \in \text{ri } \mathfrak{D}_{\mathcal{K}_Q}^{(c)}$. Then $v \in \text{aff } \mathfrak{D}_{\mathcal{K}_Q}^{(c)}$, and hence [by (ii)] $v \in \text{aff } \mathfrak{D}_M^{(c)}$ for sufficiently large M . Thus $\langle \alpha, v \rangle = 0$, all $\alpha \in (\text{aff } \mathfrak{D}_{\mathcal{K}_Q}^{(c)})^\perp$ and all $\alpha \in (\text{aff } \mathfrak{D}_M^{(c)})^\perp$, $M \geq \text{some } M_0$. Using this fact, we may then proceed as in the previous paragraph, replacing $I_{\mathcal{K}_Q}^{(c)}$, $I_M^{(c)}$ with their restrictions to $\text{aff } \mathfrak{D}_{\mathcal{K}_Q}^{(c)}$.

(iv) Let \check{v} , $\check{\Lambda}_M$, $\check{I}_M^{(c)}$ and $\check{\mathfrak{D}}_M^{(c)}$ denote the restrictions of v , Λ_M , $I_M^{(c)}$ and $\mathfrak{D}_M^{(c)}$ to $\text{aff } \mathfrak{D}_M^{(c)}$. Then $\check{v} \in \text{int } \check{\mathfrak{D}}_M^{(c)}$. Hence $\partial \check{I}_M^{(c)}(v) \neq \emptyset$ [Rockafellar (1970), Theorem 23.4]. This implies

$$(4.26) \quad \check{I}_M^{(c)}(\check{v}) \stackrel{\text{def}}{=} \sup_{\beta \in \mathcal{L}_c \check{\Lambda}_M} \langle \beta, \check{v} \rangle = \langle \check{\theta}, \check{v} \rangle \quad \text{for some } \check{\theta} \in \partial(\mathcal{L}_c \check{\Lambda}_M)$$

[Rockafellar (1970), Theorem 23.5]. Since $\langle \alpha, v \rangle = 0$ for all $\alpha \in (\text{aff } \mathfrak{D}_M^{(c)})^\perp$, (4.26) also holds with $I_M^{(c)}$ in place of $\check{I}_M^{(c)}$, etc., and θ in place of $\check{\theta}$, where $\theta = \check{\theta}$ on $\text{aff } \mathfrak{D}_M^{(c)}$ and $\theta = 0$ on $(\text{aff } \mathfrak{D}_M^{(c)})^\perp$.

Finally observe by (4.26) that \check{v} is normal to $\mathcal{L}_c \check{\Lambda}_M$ at $\check{\theta}$; hence v is normal to $\mathcal{L}_c \Lambda_M$ at θ . If $c > \inf_{\alpha} \Lambda_M(\alpha)$, then it follows from Corollary 23.7.1 of Rockafellar (1970) that $v = \rho \nabla \Lambda_M(\theta)$, for some constant $\rho \geq 0$. \square

LEMMA 4.4. *Let $\{(X_n, S_n) : n = 0, 1, \dots\}$ be a Markov additive process on $\mathbb{S} \times \mathbb{R}^1$ satisfying (\mathfrak{M}') . Let \mathcal{P} denote its transition kernel, and assume that the additive components $\{\xi_n\}$ and regeneration times $\{\tau_i\}$ are bounded and $\mathbf{E}_{\pi}(S_1) = 0$, where π is the stationary distribution of $\{X_n\}$. Then, for any $\Delta > 0$ and $K > 0$,*

$$(4.27) \quad \lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{P} \left\{ \max_{0 \leq n \leq \lfloor K/\varepsilon \rfloor} |S_n| \geq \frac{\Delta}{\varepsilon} \right\} = - \inf_{t \in (0, K/\varepsilon]} \left\{ \inf_{v = \pm \Delta} t \Lambda_{\mathcal{P}}^* \left(\frac{v}{t} \right) \right\} < 0,$$

where $(\lambda_{\mathcal{P}}(\alpha))^{-1}$ is the convergence parameter of $\hat{\mathcal{P}}(\alpha)$.

PROOF. See Collamore [(1998), Theorem 1]. Since $\Lambda_{\mathcal{P}}^*(v) = 0 \Leftrightarrow v = \mathbf{E}_{\pi}(S_1) = 0$, the right-hand side of (4.27) is less than 0.

We remark that hypothesis (H2) of Collamore (1998) is not needed when the time interval ($= [0, K]$ in this case) is bounded, and hypothesis (H0) of that paper is satisfied by the results of Ney and Nummelin (1987b). The “s-set” assumption in Theorem 2 of Ney and Nummelin (1987b) is not needed, because $\{\xi_n\}$ and $\{\tau_i\}$ are bounded; hence $r_{\mathcal{P}}(\cdot; \alpha)$ is uniformly positive for all α , by Lemma 2.2(i), and inspection of the proof shows that the “s-set” condition is unnecessary in that case. \square

Proof of Theorem 3.1 (Lower bound).

Case 1: $\mathcal{L}_0 \Lambda_{\mathcal{K}_{\mathcal{Q}}} \neq \emptyset$. Let $v \in \mathfrak{A} \cap \text{ri } \mathcal{D}_{\mathcal{K}_{\mathcal{Q}}} - \{0\}$. Then $v \in \text{ri } \mathcal{D}_M$ for sufficiently large M [Lemma 4.3(ii)]. Assume M has been chosen so that this is true. Then by Lemma 4.3(iv), there exists $\theta \in \partial(\mathcal{L}_0 \Lambda_M)$ and a positive constant ρ such that $\rho \nabla \Lambda_M(\theta) = v$.

Define

$$(4.28) \quad \mathcal{R}_{\theta}(x, dy \times ds) = \frac{e^{(\theta, s)} r_M(y; \theta)}{r_M(x; \theta)} \mathcal{K}_{\mathcal{Q}}^M(x, dy \times ds),$$

and observe that $\Lambda_M(\theta) = 0 \Rightarrow \mathcal{R}_{\theta}$ is itself a Markov additive probability kernel.

Let $\mathfrak{P}_k^{\varepsilon}$ denote the paths which first hit A/ε at time $T^{\varepsilon}(A) = k$, and let $\tilde{x}_0 = (x_0, 0)$. Then, by (3.4) and (4.18),

$$(4.29) \quad \begin{aligned} \mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2) &\geq \sum_{k=1}^{\infty} \int_{\mathfrak{P}_k^{\varepsilon}} \mathcal{K}_{\mathcal{Q}}^M(\tilde{x}_0, dx_1 \times ds_1) \cdots \mathcal{K}_{\mathcal{Q}}^M(x_{k-1}, dx_k \times ds_k) \\ &= \sum_{k=1}^{\infty} \int_{\mathfrak{P}_k^{\varepsilon}} \frac{e^{-(\theta, s_1 + \cdots + s_k)} r_M(\tilde{x}_0; \theta)}{r_M(x_k; \theta)} \\ &\quad \times \mathcal{R}_{\theta}(\tilde{x}_0, dx_1 \times ds_1) \cdots \mathcal{R}_{\theta}(x_{k-1}, dx_k \times ds_k). \end{aligned}$$

To analyze the quantity on the right-hand side, note that $\mathbf{E}_{\pi_\theta}(\xi_n) = \nabla \Lambda_M(\theta) = v/\rho$, where π_θ is the stationary distribution of $\{X_n\}$ under \mathcal{R}_θ [Ney and Nummelin (1987a), Lemmas 3.3 and 5.2]. Thus, the expected time for the \mathcal{R}_θ -process to reach the point $v/\varepsilon \in (\partial A)/\varepsilon$ is approximately ρ/ε . Also, since $v \in \mathfrak{A}$, the straight-line path $[0, v]$ contains no points *other* than v belonging to the convex set $\text{cl } A$. Hence, by Lemma 4.4,

$$\mathbf{P}_{\mathcal{R}_\theta}\{T^\varepsilon(A) \leq (\rho - \Delta)/\varepsilon\} \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0, \text{ for any } \Delta > 0;$$

in other words, the process stays near its central tendency and therefore does *not* enter A/ε before the expected time of ρ/ε . By an analogous argument, we also obtain

$$\mathbf{P}_{\mathcal{R}_\theta}\{T^\varepsilon(A) \leq \rho/\varepsilon, S_{T^\varepsilon(A)} \in B(v, \Delta)^c/\varepsilon\} \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0,$$

where $B(v, \Delta)$ is a Δ -ball about v . Finally, by the central limit theorem for Markov additive processes,

$$\liminf_{\varepsilon \rightarrow 0} \mathbf{P}_{\mathcal{R}_\theta}\{T^\varepsilon(A) \leq \rho/\varepsilon\} \geq \text{const} > 0.$$

Putting these together yields

$$(4.30) \quad \liminf_{\varepsilon \rightarrow 0} \mathbf{P}\{\varepsilon T^\varepsilon(A) \in (\rho - \Delta, \rho], \varepsilon S_{T^\varepsilon(A)} \in B(v, \Delta)\} \geq \text{const} > 0.$$

Since $r_M(\cdot; \theta)$ is positive and bounded, by Lemma 4.2(iii), it follows from (4.29) and (4.30) that

$$(4.31) \quad \liminf_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}(\mathfrak{E}_{\mathcal{Q}, \varepsilon}^2) \geq -\langle \theta, v \rangle - \Delta \|\theta\| = -I_M(v) - \Delta \|\theta\|.$$

Now let $\Delta \rightarrow 0$ and then $M \rightarrow \infty$. From Lemma 4.3(iii) and (4.31), we then obtain

$$(4.32) \quad \liminf_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}(\mathfrak{E}_{\mathcal{Q}, \varepsilon}^2) \geq -I_{\mathcal{K}_\mathcal{Q}}(v).$$

The required lower bound follows by taking the supremum in (4.32) over $v \in (\mathfrak{A} \cap \text{ri } \mathfrak{D}_{\mathcal{K}_\mathcal{Q}} - \{0\})$, and observing by Lemma 4.3(i) and the definition of \mathfrak{A} that $A \cap \text{ri } \mathfrak{D}_{\mathcal{P}} \neq \emptyset \Rightarrow \mathfrak{A} \cap \text{ri } \mathfrak{D}_{\mathcal{K}_\mathcal{Q}} \neq \emptyset$. Hence $\inf\{I_{\mathcal{K}_\mathcal{Q}}(v) : v \in \mathfrak{A} \cap \text{ri } \mathfrak{D}_{\mathcal{K}_\mathcal{Q}} - \{0\}\} = \inf\{I_{\mathcal{K}_\mathcal{Q}}(v) : v \in \mathfrak{A}\}$ [cf. Collamore (1996a), the last paragraph in the proof of Theorem 2.1].

Case 2: $\mathcal{L}_0 \Lambda_{\mathcal{K}_\mathcal{Q}} = \emptyset$. Let $M_0 = \min\{M \in \mathbb{Z}_+ : \mathcal{L}_0 \Lambda_M = \emptyset\} \leq \infty$, and first assume $M_0 < \infty$.

For each M , let $c_M = \inf_{\alpha \in \mathbb{R}^d} \Lambda_M(\alpha)$ and $c = \inf_{\alpha \in \mathbb{R}^d} \Lambda_{\mathcal{K}_\mathcal{Q}}(\alpha)$. Then $\mathfrak{D}_M^{(b_M)} \nearrow \mathfrak{D}_{\mathcal{K}_\mathcal{Q}}^{(b)}$ as $M \rightarrow \infty$, for any $b_M > c_M$ and $b > c$. Hence $A \cap \text{ri } \mathfrak{D}_{\mathcal{P}} \neq \emptyset \Rightarrow A \cap \text{ri } \mathfrak{D}_M^{(b_M)} \neq \emptyset$ ($b_M > c_M$) for sufficiently large M (Lemma 4.3).

Let $M \geq M_0$ be chosen such that $A \cap \text{ri } \mathfrak{D}_M^{(b_M)} \neq \emptyset$ ($b_M > c_M$), and let $v \in \mathfrak{A} \cap \text{ri } \mathfrak{D}_M^{(b_M)} - \{0\}$. Let $d_j = c_M + 1/j > 0$, $j = 1, 2, \dots$. Then by Lemma 4.3(iv),

there exist elements $\theta_j \in \partial(\mathcal{L}_j \Lambda_M) \cap \text{aff } \mathcal{D}_M^{(d_j)}$ and positive constants ρ_j such that $\rho_j \nabla \Lambda_M(\theta_j) = v$.

For each j , introduce the Markov additive probability kernel

$$(4.33) \quad \mathcal{R}_{\theta_j}(x, dy \times ds) = \frac{e^{(\theta_j, s) - \Lambda_M(\theta_j)} r_M(y; \theta_j)}{r_M(x; \theta_j)} \mathcal{K}_{\mathcal{Q}}^M(x, dy \times ds),$$

and reason as in (4.29) to obtain

$$(4.34) \quad \mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2) \geq \sum_{k=1}^{\infty} \int_{\mathfrak{P}_k^s} \frac{e^{-(\theta_j, s_1 + \dots + s_k) + k \Lambda_M(\theta_j)} r_M(x_0; \theta_j)}{r_M(x_k; \theta_j)} \times \mathcal{R}_{\theta_j}(x_0, dx_1 \times ds_1) \cdots \mathcal{R}_{\theta_j}(x_{k-1}, dx_k \times ds_k).$$

It follows from (4.30) and (4.34) that

$$(4.35) \quad \liminf_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2) \geq -\langle \theta_j, v \rangle + \rho_j \Lambda_M(\theta_j).$$

We now distinguish two possible cases. First, suppose that $\{\theta_j\}$ converges (possibly after passing to a subsequence) to some element $\hat{\theta} \in \mathbb{R}^d$. Then the infimum in the definition of c_M is *achieved* at $\hat{\theta}$; hence $\Lambda_M(\hat{\theta}) = c_M > 0$ and $\nabla \Lambda_M(\hat{\theta}) = 0$. But then $\lim_{j \rightarrow \infty} \rho_j = \lim_{j \rightarrow \infty} (v / \nabla \Lambda_M(\theta_j)) = \infty$. Letting $j \rightarrow \infty$ in (4.35), we conclude

$$(4.36) \quad \lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2) = \infty.$$

Next, suppose that $\{\theta_j\}$ does not converge along any subsequence. Let $\beta_j = \theta_j / \|\theta_j\|$, and observe that (possibly after passing to a subsequence) $\beta_j \rightarrow \hat{\beta} \in \mathbf{S}^{d-1}$ and $\|\theta_j\| \rightarrow \infty$ as $j \rightarrow \infty$. Then $\hat{\beta} \in 0^+ \Lambda_M$ [Rockafellar (1970), Theorems 8.2 and 8.7]. Hence $\hat{\beta} \in (\mathcal{D}_M^{(c_M+1)})^\circ$ [as in the proof of Lemma 4.3(i)]. Since the θ_j 's were chosen in $\text{aff } \mathcal{D}_M^{(d_j)} = \text{aff } \mathcal{D}_M^{(c_M+1)}$, it follows that $\hat{\beta} \in (\mathcal{D}_M^{(c_M+1)})^\circ \cap \text{aff } \mathcal{D}_M^{(c_M+1)}$. Then $v \in \text{ri } \mathcal{D}_M^{(c_M+1)} \Rightarrow \langle \hat{\beta}, v \rangle < 0$. Hence $\langle \beta_j, v \rangle \leq -a_0 < 0$, for all $j \geq \text{some } j_0$. But then

$$(4.37) \quad -\langle \theta_j, v \rangle = -\|\theta_j\| \langle \beta_j, v \rangle \rightarrow \infty \quad \text{as } j \rightarrow \infty.$$

Thus, letting $j \rightarrow \infty$ in (4.35) we again obtain (4.36).

Finally suppose $M_0 = \infty$. In this case, the elements of $\{\mathcal{L}_0 \Lambda_M : M = 1, 2, \dots\}$ are nonempty and monotonically decreasing to $\bigcap_M \mathcal{L}_0 \Lambda_M = \mathcal{L}_0 \Lambda_{\mathcal{K}_{\mathcal{Q}}} = \emptyset$. Then

$$(4.38) \quad \inf\{\|\alpha\| : \alpha \in \mathcal{L}_0 \Lambda_M\} \rightarrow \infty \quad \text{as } M \rightarrow \infty.$$

Note that Λ_M is strictly increasing, which (with $M_0 = \infty$) implies $\inf_{\alpha} \Lambda_M(\alpha) < 0, \forall M$. Since $A \cap \text{ri } \mathcal{D}_{\mathcal{P}} \neq \emptyset$ and \mathcal{D}_M increases as $M \rightarrow \infty$, Lemma 4.3(i), (ii) implies the existence of an element $v \in \bigcap_{M \geq M_0'} (\mathcal{A} \cap \text{ri } \mathcal{D}_M - \{0\}) \subset \text{ri } \mathcal{D}_{\mathcal{K}_{\mathcal{Q}}}^{(1)}$. Applying Lemma 4.3(iv), we then obtain elements $\theta_M \in (\partial(\mathcal{L}_0 \Lambda_M) \cap \text{aff } \mathcal{D}_M)$

and positive constants ρ_M such that $\rho_M \nabla \Lambda_M(\theta_M) = v$. By (4.38), $\|\theta_M\| \rightarrow \infty$ as $M \rightarrow \infty$.

Let $\beta_M = \theta_M / \|\theta_M\|$. Then (possibly after passing to a subsequence) $\beta_M \rightarrow \hat{\beta} \in \bigcap_M 0^+ \Lambda_M$ [Rockafellar (1970), Theorems 8.2 and 8.7]. Then $\hat{\beta} \in (\mathfrak{D}_{\mathcal{K}_Q}^{(1)})^\circ$ [Lemma 4.3(ii) and its proof]. Then $v \in (\mathfrak{D}_{\mathcal{K}_Q}^{(1)})^\circ$ and $v \in \text{ri } \mathfrak{D}_{\mathcal{K}_Q}^{(1)} \Rightarrow \langle \hat{\beta}, v \rangle < 0$. Hence $\langle \beta_M, v \rangle \leq -a'_0 < 0$, for M sufficiently large. Hence (4.35) and (4.37) (with “ M ” in place of “ j ”) give (4.36), as before.

4.3. *Proofs of Theorem 3.4 and Proposition 3.5.* Next we turn to the proof of Theorem 3.4. Let \mathcal{Q}^* be the kernel described in (3.18), and let $\mathcal{Q} \in \mathcal{C}$. By the Radon–Nikodym theorem we may write $\mathcal{Q}^{(t)}(x, \cdot) = \mathcal{R}^{(t)}(x, \cdot) + \mathcal{V}^{(t)}(x, \cdot)$, where $\mathcal{R}^{(t)}(x, \cdot) \ll \mathcal{Q}^*(x, \cdot)$ and $\mathcal{V}^{(t)}(x, \cdot) \perp \mathcal{Q}^*(x, \cdot)$, for any given $x \in \mathbb{S}$. Define

$$(4.39) \quad Z_{n+1}^\varepsilon = \log \left(\frac{d\mathcal{R}^{(\varepsilon n)}}{d\mathcal{Q}^*}(X_n^*, X_{n+1}^* \times \xi_{n+1}^*) \right), \quad n = 0, 1, \dots,$$

where $\{(X_n^*, S_n^*) : n = 0, 1, \dots\}$ denotes a Markov additive process with transition kernel \mathcal{Q}^* . Let

$$W_n^\varepsilon = Z_1^\varepsilon + \dots + Z_n^\varepsilon, \quad n = 1, 2, \dots, \text{ and } W_0^\varepsilon = 0.$$

The proof of Theorem 3.4 will rely on the following.

LEMMA 4.5. (i) *For any fixed ε , $\{W_n^\varepsilon\}$ is a submartingale.*

(ii) *If $\check{Z}_n^\varepsilon = Z_n^\varepsilon \vee 0 - 1$ and $\check{W}_n^\varepsilon = \check{Z}_1^\varepsilon + \dots + \check{Z}_n^\varepsilon$, $n \in \mathbb{Z}_+$, then $\{\check{W}_n^\varepsilon\}$ is a submartingale.*

(iii) *Suppose $\mathcal{Q} \in \mathcal{C}_0$, so that $\{Z_n^\varepsilon\}$, $\{W_n^\varepsilon\}$ are actually independent of ε , and let $Z_n^M = Z_n^\varepsilon \vee (-M)$, $W_n^M = Z_1^M + \dots + Z_n^M$, $n \in \mathbb{Z}_+$, and $W_0^M = 0$. Assume that the lower bound of (\mathfrak{X}) holds, and assume that we do not have $\mathcal{Q} = \mathcal{Q}^*$ for \mathcal{P} a.e. (y, s) , φ a.e. x . Then, for some positive constant D ,*

$$(4.40) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{E}_{\mathcal{Q}^*}(W_n^M) \leq -D \quad \text{for all } M \geq \text{some } M_0.$$

PROOF. (i) Jensen’s inequality implies that $\mathbf{E}(Z_{n+1}^\varepsilon \mid X_n^* = x) \leq 0$ for all x ; hence $\{W_n^\varepsilon\}$ is a submartingale.

(ii) This follows by a similar argument and the inequality $(\log s) \leq s$.

(iii) Let $\{(\tilde{X}_n^*, \tilde{S}_n^*, \tilde{W}_n) : n = 0, 1, \dots\}$ be an independent copy of $\{(X_n^*, S_n^*, W_n) : n = 0, 1, \dots\}$, but assume that the initial measure of \tilde{X}_n^* is

π^* = the stationary measure of $\{X_n^*\}$ under the transition kernel \mathcal{Q}^* .

Let $\{T_i\}_{i \in \mathbb{N}}$ and $\{\tilde{T}_i\}_{i \in \mathbb{N}}$ denote the respective regeneration times, as described in Lemma 2.1, and let

$$\mathfrak{T} \stackrel{\text{def}}{=} \inf\{n : T_i = n \text{ and } \tilde{T}_j = n, \text{ some } i, j \in \mathbb{N}\}$$

denote the coupling time.

First note that if we do not have $\mathcal{Q} = \mathcal{Q}^*$ for \mathcal{P} a.e. (y, s) , φ a.e. x , then Jensen's inequality implies $\mathbf{E}_{\pi^*}(Z_n^\varepsilon) < 0$. By the monotone convergence $Z_n^M \searrow Z_n^\varepsilon$ as $M \rightarrow \infty$, it follows that

$$(4.41) \quad \mathbf{E}_{\pi^*}(Z_n^M) \leq -D < 0 \quad \text{for all } M \geq \text{some } M_0.$$

Consequently

$$(4.42) \quad \mathbf{E}(\tilde{W}_n^M) \leq -nD \quad \text{for all } M \geq \text{some } M_0.$$

Let $\mathfrak{T}_n = \mathfrak{T} \wedge n$, and observe

$$(4.43) \quad \mathbf{E}(W_n^M) = \mathbf{E}(\tilde{W}_n^M) + \mathbf{E}(W_{\mathfrak{T}_n}^M - \tilde{W}_{\mathfrak{T}_n}^M).$$

Also, by a slight variant of (i) and (ii), $\{W_n^M - n\varepsilon_M\}$ is a submartingale for $\varepsilon_M = \log(1 + e^{-M})$. Hence by the optional sampling theorem,

$$(4.44) \quad \lim_{n \rightarrow \infty} \mathbf{E}(W_{\mathfrak{T}_n}^M) \leq \varepsilon_M \mathbf{E}(\mathfrak{T}) < \infty.$$

Since $\tilde{W}_n^M \geq -M$, we also have

$$(4.45) \quad \limsup_{n \rightarrow \infty} \{-\mathbf{E}(\tilde{W}_{\mathfrak{T}_n}^M)\} \leq M \mathbf{E}(\mathfrak{T}) < \infty.$$

[By Remark 2.1(i), $\mathbf{E}(\mathfrak{T}) < \infty$.] The required result is then obtained by substituting (4.42), (4.44) and (4.45) into (4.43). \square

LEMMA 4.6. *Let $A \subset \mathbb{R}^d$ be a convex semicone intersecting $\text{ri } \mathfrak{S}$. Assume that $\text{dom } \Lambda_{\mathcal{P}}$ is open and that (H2) and (\mathfrak{R}) are satisfied. Let ϱ be given as in Lemma 3.2, and let $\tau \stackrel{d}{=} \tau_i$, where $\tau_i = T_{i+1} - T_i$ are the interregeneration times described in Lemma 2.1. Define $\mathfrak{I}^\varepsilon(A) = \inf\{i : T_i \geq T^\varepsilon(A)\}$. Then, for φ a.e. x_0 ,*

$$(4.46) \quad \lim_{\varepsilon \rightarrow 0} \varepsilon \mathbf{E}_{\mathcal{Q}^*}(\mathfrak{I}^\varepsilon(A)) = \frac{\varrho}{\mathbf{E}_{\mathcal{Q}^*}(\tau)}.$$

PROOF. Let α_0, v_0 and a be given as in Lemma 3.2.

Lower bound. First introduce a truncation on the additive components; namely let $M > 0$ and define

$$(4.47) \quad \xi_n^M = \xi_n^* \quad \text{if } \langle \alpha_0, \xi_n^* \rangle \geq -M, \quad \text{and} \quad \xi_n^M = \xi_n^* \frac{M}{\|\langle \alpha_0, \xi_n^* \rangle\|} \quad \text{otherwise.}$$

Let $S_n^M = \xi_1^M + \cdots + \xi_n^M$, $n = 0, 1, \dots$, and $S_0^M = 0$, and let

$$B = \mathcal{H}(\alpha_0, a) = \{v : \langle \alpha_0, v \rangle > a\}.$$

Then by Lemma 3.2(i), $\mathfrak{I}^\varepsilon(A) \geq \mathfrak{I}^{M,\varepsilon}(B)$, where $\mathfrak{I}^{M,\varepsilon}(\cdot)$ denotes the stopping time with respect to the truncated process $\{S_n^M\}_{n \in \mathbb{N}}$.

Note that $\text{dom } \Lambda_{\mathcal{P}} \text{ open} \Rightarrow 0 \in \text{int}(\text{dom } \Lambda_{\mathcal{Q}^*})$. Hence we may apply the optional sampling theorem to obtain

$$(4.48) \quad \begin{aligned} \mathbf{E}_{\mathcal{Q}^*}(\langle \alpha_0, S_{T_{\mathcal{I}^{M,\varepsilon}(B)}}^M \rangle) &= \mathbf{E}_{\mathcal{Q}^*}(\langle \alpha_0, S_{T_0}^M \rangle) \\ &+ \mathbf{E}_{\mathcal{Q}^*}(\langle \alpha_0, S_{T_{i+1}}^M - S_{T_i}^M \rangle) \mathbf{E}_{\mathcal{Q}^*}(\langle \mathcal{I}^{M,\varepsilon}(B) \rangle - 1). \end{aligned}$$

Also, under the above truncation,

$$(4.49) \quad \begin{aligned} \mathbf{E}_{\mathcal{Q}^*}(\langle \alpha_0, S_{T_{\mathcal{I}^{M,\varepsilon}(B)}}^M - S_{T_{M,\varepsilon}(B)}}^M \rangle) &\geq -M \mathbf{E}_{\mathcal{Q}^*}(T_{\mathcal{I}^{M,\varepsilon}(B)} - T^{M,\varepsilon}(B)) \\ &\geq -MC, \end{aligned}$$

where [by Remark 2.1(i)] the constant $C < \infty$. Since $\langle \alpha_0, S_{T_{M,\varepsilon}(B)}}^M \rangle > a/\varepsilon$, it follows from (4.48) and (4.49) that

$$(4.50) \quad \liminf_{\varepsilon \rightarrow 0} \varepsilon \mathbf{E}_{\mathcal{Q}^*}(\langle \mathcal{I}^{M,\varepsilon}(B) \rangle) \geq a (\mathbf{E}_{\mathcal{Q}^*}(\langle \alpha_0, S_{T_{i+1}}^M - S_{T_i}^M \rangle))^{-1}, \quad \varphi \text{ a.e. } x_0,$$

provided that

$$(4.51) \quad \mathbf{E}_{\mathcal{Q}^*}(\langle \alpha_0, S_{T_0}^M \rangle) < \infty, \quad \varphi \text{ a.e. } x_0.$$

Finally observe by Ney and Nummelin [(1987a), Lemma 3.3] and Lemma 3.2(v),

$$(4.52) \quad \mathbf{E}_{\mathcal{Q}^*}(S_{T_{i+1}}^* - S_{T_i}^*) = \nabla \Lambda_{\mathcal{P}}(\alpha_0) \cdot \mathbf{E}_{\mathcal{Q}^*}(\tau) = \frac{v_0}{\varrho} \mathbf{E}_{\mathcal{Q}^*}(\tau), \quad i = 0, 1, \dots$$

Then (4.51) holds, and the required result follows from (4.50), (4.52), Lemma 3.2(iii) and the monotone convergence $\langle \alpha_0, S_{T_{i+1}}^M - S_{T_i}^M \rangle \searrow \langle \alpha_0, S_{T_{i+1}}^* - S_{T_i}^* \rangle$ as $M \rightarrow \infty$.

Upper bound. First assume $d > 1$. Let $t > 0$, and observe that since A is a semicone, $(1+t)v_0$ is an interior point of A . Choose $w^{(1)}, \dots, w^{(d)} \in A$ such that the convex hull of $\{v_0, w^{(1)}, \dots, w^{(d)}\}$ contains a neighborhood of $(1+t)v_0$. Let $v^{(k)} = v_0 + w^{(k)}$; let $\mathcal{V}^{(k)}$ be the hyperplane containing $\{v_0, v^{(1)}, \dots, v^{(d)}\} - \{v^{(k)}\}$; and let $\mathcal{H}^{(k)}$ be the open half-space determined by $\mathcal{V}^{(k)}$ which contains the point $(1+t)v_0$. Then $\text{cl}(\bigcap_{k=1}^d \mathcal{H}^{(k)}) = v_0 + \text{cone}(\text{conv}\{w^{(1)}, \dots, w^{(d)}\})$ [Rockafellar (1970), Theorem 18.8]. Hence, it follows from our construction and the semicone property that $\bigcap_{k=1}^d \mathcal{H}^{(k)} \subset A$. Also, $0 \notin \mathcal{H}^{(k)}$ and $v_0 \in \partial \mathcal{H}^{(k)}$, for all k .

Let $\mathcal{I}_j^\varepsilon = \inf\{i : S_{T_i}^* \in \mathcal{H}^{(k)}, \text{ all } j \geq i\}$. By (4.52), the expected time for $\{S_{T_i}^*\}_{i \in \mathbb{N}}$ to reach $\mathcal{H}^{(k)}$ is $i = \varrho/\mathbf{E}(\tau)$. Hence, by a simple one-dimensional change of measure argument,

$$(4.53) \quad \lim_{\varepsilon \rightarrow 0} \mathbf{E}_{\mathcal{Q}^*} \left[\mathcal{I}_j^\varepsilon; \mathcal{I}_j^\varepsilon \geq \frac{1}{\varepsilon} \left(\frac{\varrho + \Delta}{\mathbf{E}_{\mathcal{Q}^*}(\tau)} \right) \right] = 0 \quad \text{for all } \Delta > 0$$

[cf. Collamore (1998), Theorem 4.1, for a closely related result]. Since $\mathcal{I}^\varepsilon(A) \leq \max\{\mathcal{I}_1^\varepsilon, \dots, \mathcal{I}_d^\varepsilon\}$, the upper bound is obtained from (4.53).

Finally, if $d = 1$, then the upper bound can be obtained directly from (4.53). \square

PROOF OF THEOREM 3.4. Following Asmussen and Rubinstein [(1995), Theorem 17.6], first observe that

$$(4.54) \quad d\mathcal{K}_{\mathcal{Q}}^{(\varepsilon n)} \stackrel{\text{def}}{=} \left(\frac{d\mathcal{P}}{d\mathcal{Q}^{(\varepsilon n)}} \right)^2 d\mathcal{Q}^{(\varepsilon n)} = \left(\frac{d\mathcal{P}}{d\mathcal{Q}^*} \right)^2 \left(\frac{d\mathcal{Q}^*}{d\mathcal{Q}^{(\varepsilon n)}} \right) d\mathcal{Q}^*.$$

Also, by the Radon–Nikodym theorem and the definition of $\mathcal{R}^{(\varepsilon n)}$,

$$(4.55) \quad \frac{d\mathcal{Q}^*}{d\mathcal{Q}^{(\varepsilon n)}} = \frac{d\mathcal{Q}^*}{d\mathcal{R}^{(\varepsilon n)}} = \left(\frac{d\mathcal{R}^{(\varepsilon n)}}{d\mathcal{Q}^*} \right)^{-1}, \quad \mathcal{Q}^* \text{ a.e.}$$

From (4.54) and (4.55) it follows that

$$(4.56) \quad \begin{aligned} \mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2) &= \mathbf{E}_{\mathcal{Q}^*} \left[\frac{r_{\mathcal{P}}^2(x_0; \alpha_0)}{r_{\mathcal{P}}^2(X_{T^\varepsilon(A)}; \alpha_0)} \exp(-2\langle \alpha_0, S_{T^\varepsilon(A)}^* \rangle - W_{T^\varepsilon(A)}^\varepsilon) \right] \\ &\geq \exp\{-2\mathbf{E}_{\mathcal{Q}^*}(\langle \alpha_0, S_{T^\varepsilon(A)}^* \rangle) - \mathbf{E}_{\mathcal{Q}^*}(W_{T^\varepsilon(A)}^\varepsilon) + C'\}, \end{aligned}$$

where the last step was obtained by Jensen's inequality, and C' is a finite constant obtained from the uniform positivity and boundedness of $r_{\mathcal{P}}(\alpha_0)$ described in Lemma 2.2(iii).

Let T_0, T_1, \dots denote the regeneration times in Lemma 2.1 generated by the Markov additive process $\{(X_n, S_n)\}$; let $\{\tau_i\}$ denote the interregeneration times; and let

$$\mathcal{I}^\varepsilon(A) = \inf\{i : T_i \geq T^\varepsilon(A)\}.$$

Introduce the truncation $\{(\xi_n^M, S_n^M) : n = 0, 1, \dots\}$ of $\{(\xi_n^*, S_n^*) : n = 0, 1, \dots\}$ that was described above in (4.47), and observe under this truncation that (4.49) holds with A in place of B and $\mathcal{I}^\varepsilon(\cdot)$ in place of $\mathcal{I}^{M, \varepsilon}(\cdot)$. Hence, it follows from (4.56) and the definition of $\{S_n^M\}$ that

$$(4.57) \quad \log \mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2) \geq -2\mathbf{E}_{\mathcal{Q}^*}(\langle \alpha_0, S_{T_{\mathcal{I}^\varepsilon(A)}^M}^* \rangle) - \mathbf{E}_{\mathcal{Q}^*}(W_{T^\varepsilon(A)}^\varepsilon) + C$$

for some constant $C \in (-\infty, \infty)$.

By the optional sampling theorem and Lemma 4.5(i), (ii),

$$(4.58) \quad \mathbf{E}_{\mathcal{Q}^*}(W_{T^\varepsilon(A)}^\varepsilon) \leq 0.$$

Also, by the optional sampling theorem,

$$(4.59) \quad \begin{aligned} \mathbf{E}_{\mathcal{Q}^*}(\langle \alpha_0, S_{T_{\mathcal{I}^\varepsilon(A)}^M}^* \rangle) &= \mathbf{E}_{\mathcal{Q}^*}(\langle \alpha_0, S_{T_0}^M \rangle) \\ &\quad + \mathbf{E}_{\mathcal{Q}^*}(\langle \alpha_0, S_{T_{i+1}}^M - S_{T_i}^M \rangle) \mathbf{E}_{\mathcal{Q}^*}((T_{\mathcal{I}^\varepsilon(A)} - 1)). \end{aligned}$$

Then by (4.51), (4.52), Lemma 4.6 and the monotone convergence $\langle \alpha_0, S_{T_{i+1}}^M - S_{T_i}^M \rangle \searrow \langle \alpha_0, S_{T_{i+1}}^* - S_{T_i}^* \rangle$,

$$(4.60) \quad \lim_{\varepsilon \rightarrow 0} \varepsilon \mathbf{E}_{\mathcal{Q}^*}(\langle \alpha_0, S_{T_{\mathcal{I}^\varepsilon(A)}^M}^* \rangle) \searrow \langle \alpha_0, v_0 \rangle \quad \text{as } M \rightarrow \infty.$$

From (4.57), (4.58) and (4.60) we conclude

$$(4.61) \quad \liminf_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2) \geq -2\langle \alpha_0, v_0 \rangle.$$

Together with Lemma 3.2(iii) and Theorem 3.3(ii), this implies

$$(4.62) \quad \liminf_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}(\mathcal{E}_{\mathcal{Q}, \varepsilon}^2) \geq \lim_{\varepsilon \rightarrow 0} \varepsilon \log \mathbf{E}(\mathcal{E}_{\mathcal{Q}^*, \varepsilon}^2).$$

It remains to show that if $\mathcal{Q}^{(t_0)} \neq \mathcal{Q}^*$ at some continuity point $t_0 \in [0, \varrho]$, then there is strict inequality in (4.62). Suppose now that $t_0 \in [0, \varrho]$ is a continuity point and $\mathcal{Q}^{(t_0)} \neq \mathcal{Q}^*$. Then the continuity properties (3.25) and (3.26) are satisfied in some interval $[\zeta_1, \zeta_2] \subset [0, \varrho]$. Let D and M_0 be the constants obtained in Lemma 4.5(iii) when $\mathcal{Q}_0 = \mathcal{Q}^{(t_0)}$. Assume that the interval $[\zeta_1, \zeta_2]$ has been chosen sufficiently small that (3.25) holds with $\Delta = D/2$.

Decompose the random variable Z_n^ε into a sum of two terms, namely,

$$U_n = \log \left(\frac{d\mathcal{R}^{(t_0)}}{d\mathcal{Q}^*}(X_n^*, X_{n+1}^* \times \xi_{n+1}^*) \right), \quad n = 0, 1, \dots,$$

$$V_n^\varepsilon = \log \left(\frac{d\mathcal{R}^{(\varepsilon n)}}{d\mathcal{R}^{(t_0)}}(X_n^*, X_{n+1}^* \times \xi_{n+1}^*) \right), \quad n = 0, 1, \dots$$

For $M > 0$, let $U_n^M = U_n \vee (-M)$; $V_n^{M, \varepsilon} = V_n^\varepsilon \vee (-M)$;

$$R_n^M = \sum_{i=0}^n U_i^M; \quad Z_n^{M, \varepsilon} = U_n^M + V_n^{M, \varepsilon}; \quad W_n^{M, \varepsilon} = \sum_{i=0}^n Z_i^{M, \varepsilon}.$$

By Lemma 4.5(iii),

$$(4.63) \quad \limsup_{\varepsilon \rightarrow 0} \varepsilon \mathbf{E}_{\mathcal{Q}^*} \left(R_{\lfloor \zeta_2 / \varepsilon \rfloor}^M - R_{\lfloor \zeta_1 / \varepsilon \rfloor}^M \right) \leq - \left(\frac{\zeta_2 - \zeta_1}{\zeta_2} \right) D \quad \text{for all } M \geq M_0.$$

Hence, from the continuity properties (3.25) and (3.26) and a straightforward variant of Lemma 4.5(iii) (applied to $\{V_n^{M, \varepsilon}\}$), we obtain

$$(4.64) \quad \limsup_{\varepsilon \rightarrow 0} \varepsilon \mathbf{E}_{\mathcal{Q}^*} \left(W_{\lfloor \zeta_2 / \varepsilon \rfloor}^{M, \varepsilon} - W_{\lfloor \zeta_1 / \varepsilon \rfloor}^{M, \varepsilon} \right) \leq -D' < 0 \quad \text{for all } M \geq M_0.$$

Moreover, by the optional sampling theorem and Lemma 4.5(i), (ii),

$$(4.65) \quad \mathbf{E}_{\mathcal{Q}^*} \left(W_{T^\varepsilon(A) \wedge \lfloor \zeta_1 / \varepsilon \rfloor}^\varepsilon \right) \leq 0$$

and

$$(4.66) \quad \mathbf{E}_{\mathcal{Q}^*} \left(W_{T^\varepsilon(A)}^\varepsilon - W_{\lfloor \zeta_2 / \varepsilon \rfloor}^\varepsilon; T^\varepsilon(A) > \left\lfloor \frac{\zeta_2}{\varepsilon} \right\rfloor \right) \leq 0.$$

It follows from (4.64)–(4.66) and the definition of $\{W_n^{M,\varepsilon}\}$ that

$$(4.67) \quad \begin{aligned} & \limsup_{\varepsilon \rightarrow 0} \varepsilon \mathbf{E}_{\mathcal{Q}^*}(W_{T^\varepsilon(A)}^\varepsilon) \\ & \leq -D' - \liminf_{\varepsilon \rightarrow 0} \varepsilon \mathbf{E}_{\mathcal{Q}^*} \left(W_{\lfloor \zeta_2/\varepsilon \rfloor}^{M,\varepsilon} - W_{T^\varepsilon(A) \wedge \lfloor \zeta_1/\varepsilon \rfloor}^{M,\varepsilon}; T^\varepsilon(A) < \frac{\zeta_2}{\varepsilon} \right). \end{aligned}$$

Since $\zeta_2 < \varrho$, $\mathbf{P}\{T^\varepsilon(A) < \lfloor \zeta_2/\varepsilon \rfloor\} \xrightarrow{\text{exp}} 0$ [Collamore (1998), Theorem 1; cf. (4.52) and the proof of Lemma 4.4]. Also, the above definitions imply that the last integrand in (4.67) is bounded below by $-2M[(\zeta_2 - \zeta_1)/\varepsilon + 1]$. We conclude that the last term on the right-hand side of (4.67) can actually be dropped.

Using (4.67) in place of (4.58) now gives strict inequality in (4.62), as desired. \square

PROOF OF PROPOSITION 3.5. (i) By definition,

$$(\mathcal{L}_b I_{\mathcal{P}})^c = \bigcup_{\alpha \in \mathcal{L}_0 \Lambda_{\mathcal{P}}} \mathcal{H}(\alpha, b) \quad \text{for all } b \geq 0.$$

Hence $\{\mathcal{H}(\alpha, a - \Delta)\}_{\alpha \in \mathcal{L}_0 \Lambda_{\mathcal{P}}}$ is an open cover for $\mathfrak{B} \stackrel{\text{def}}{=} \partial(\mathcal{L}_a I_{\mathcal{P}}) \cap (\text{cone}_\delta(\mathcal{L}_0 \Lambda_{\mathcal{P}}^*))^c$.

The set \mathfrak{B} is compact, since $I_{\mathcal{P}}$ is positively homogeneous and strictly positive on the compact set $\mathbf{S}^{d-1} \cap (\text{cone}_\delta(\mathcal{L}_0 \Lambda_{\mathcal{P}}^*))^c$ [Collamore (1996b), Lemma 3.1]. Hence \mathfrak{B} has a finite subcover. This subcover also covers $(\mathcal{L}_a I_{\mathcal{P}})^c \cap (\text{cone}_\delta(\mathcal{L}_0 \Lambda_{\mathcal{P}}^*))^c$, and hence A .

(ii) This is established in the same way as the upper bound of Theorem 3.1, with α_i in place of θ . In the case where A is a finite union of convex sets, choose the α_i 's to be the elements obtained in Lemma 3.2 when $A = A'_i$, $i = 1, \dots, k$, and then proceed as in the proof of the upper bound of Theorem 3.1. \square

Acknowledgments. This work was begun while the author was a postdoctoral research associate at the Centre for Mathematical Sciences, Lund University, Sweden, and completed while the author was a postdoctoral research associate at EURANDOM, Eindhoven University of Technology, The Netherlands. The author is grateful to these centers for providing a hospitable and stimulating research environment.

REFERENCES

- ASMUSSEN, S. (1989). Risk theory in a Markovian environment. *Scand. Actuar. J.* 69–100.
- ASMUSSEN, S. (1999). Stochastic simulation with a view towards stochastic processes. Univ. Aarhus.
- ASMUSSEN, S. and RUBINSTEIN, R. Y. (1995). Steady-state rare events simulation in queueing models and its complexity properties. In *Advances in Queueing: Models, Methods and Problems* (J. Dshalalow, ed.) 429–461. CRC Press, Boca Raton, FL.

- ATHREYA, K. B. and NEY, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245** 493–501.
- BOROVKOV, A. A. (1997). Limit theorems for time and place of the first boundary passage by a multidimensional random walk. *Doklady* **55** 254–256.
- BOROVKOV, A. A. and MOGULSKII, A. A. (1996). Large deviations for stationary Markov chains in a quarter plane. In *Probability Theory and Mathematical Statistics*. World Scientific, River Edge, NJ.
- BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York.
- BUCKLEW, J. (1998). The blind simulation problem and regenerative processes. *IEEE Trans. Inform. Theory* **44** 2877–2891.
- BUCKLEW, J. A., NEY, P. and SADOWSKY, J. S. (1990). Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains. *J. Appl. Probab.* **27** 44–59.
- CHEN, J.-C., LU, D., SADOWSKY, J. S. and YAO, K. (1993). On importance sampling in digital communications. I. Fundamentals. II. Trellis-coded modulation. *IEEE J. Selected Areas Comm.* **11** 289–308.
- COLLAMORE, J. F. (1996a). Hitting probabilities and large deviations. *Ann. Probab.* **24** 2065–2078.
- COLLAMORE, J. F. (1996b). Large deviation techniques for the study of the hitting probabilities of rare sets. Ph.D. dissertation, Univ. Wisconsin–Madison.
- COLLAMORE, J. F. (1998). First passage times of general sequences of random vectors: a large deviations approach. *Stochastic Process. Appl.* **78** 97–130.
- COTTRELL, M., FORT, J.-C. and MALGOUYRES, G. (1983). Large deviations and rare events in the study of stochastic algorithms. *IEEE Trans. Automat. Control* **28** 907–920.
- CRAMÉR, H. (1930). On the mathematical theory of risk. In *Scandinavian Jubilee Volume*. Stockholm.
- DE ACOSTA, A. (1988). Large deviations for vector-valued functionals of a Markov chain: lower bounds. *Ann. Probab.* **16** 925–960.
- DE ACOSTA, A. and NEY, P. (1998). Large deviation lower bounds for arbitrary additive functionals of a Markov chain. *Ann. Probab.* **26** 1660–1682.
- DEMO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. Springer, New York.
- DONSKER, M. D. and VARADHAN, S. R. S. (1975, 1976, 1983). Asymptotic evaluation of certain Markov process expectations for large time, I, II, III, IV. *Comm. Pure Appl. Math.* **28** 1–47; **28** 279–301; **29** 389–461; **36** 183–212.
- FREIDLIN, M. I. and WENTZELL, A. D. (1984). *Random Perturbations of Dynamical Systems*. Springer, New York.
- GLASSERMAN, P. and KOU, S. (1995). Analysis of an importance sampling estimator for tandem queues. *ACM Trans. Modeling Comp. Simul.* **4** 22–42.
- GLASSERMAN, P. and WANG, Y. (1997). Counterexamples in importance sampling for large deviations probabilities. *Ann. Appl. Probab.* **7** 731–746.
- HAMMERSLEY, J. M. and HANDSCOMB, D. C. (1964). *Monte Carlo Methods*. Chapman and Hall, London.
- ISCOE, I., NEY, P. and NUMMELIN, E. (1985). Large deviations of uniformly recurrent Markov additive processes. *Adv. in Appl. Math.* **6** 373–412.
- LEHTONEN, T. and NYRHINEN, H. (1992a). On asymptotically efficient simulation of ruin probabilities in a Markovian environment. *Scand. Actuar. J.* 60–75.
- LEHTONEN, T. and NYRHINEN, H. (1992b). Simulating level-crossing probabilities by importance sampling. *Adv. Appl. Probab.* **24** 858–874.
- LUNDBERG, F. (1909). Zur Theorie der Rückversicherung. In *Verhandl. Kongr. Versicherungsmath.* Wien.

- MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London.
- NEY, P. (1983). Dominating points and the asymptotics of large deviations for random walk on \mathbb{R}^d . *Ann. Probab.* **11** 158–167.
- NEY, P. and NUMMELIN, E. (1984). Some limit theorems for Markov additive processes. In *Proceedings of the Symposium on Semi-Markov Processes, Brussels*.
- NEY, P. and NUMMELIN, E. (1987a). Markov additive processes. I. Eigenvalue properties and limit theorems. *Ann. Probab.* **15** 561–592.
- NEY, P. and NUMMELIN, E. (1987b). Markov additive processes. II. Large deviations. *Ann. Probab.* **15** 593–609.
- NUMMELIN, E. (1978). A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete* **43** 309–318.
- NUMMELIN, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge Univ. Press.
- REVUZ, D. (1975). *Markov Chains*. North-Holland, Amsterdam.
- ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton Univ. Press.
- SADOWSKY, J. S. (1993). On the optimality and stability of exponential twisting in Monte Carlo estimation. *IEEE Trans. Inform. Theory* **39** 119–128.
- SADOWSKY, J. S. (1996). On Monte Carlo estimation of large deviations probabilities. *Ann. Appl. Probab.* **6** 399–422.
- SIEGMUND, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *Ann. Statist.* **4** 673–684.

DEPARTMENT OF MATHEMATICS
ETH ZENTRUM, HG F 42.2
CH-8092 ZÜRICH
SWITZERLAND
E-MAIL: collamore@math.ethz.ch