

Large Deviation Asymptotics and Control Variates For Simulating Large Functions*

Sean Meyn[†]

October 8, 2004

Abstract

Consider the normalized partial sums of a real-valued function F of a Markov chain,

$$\phi_n := n^{-1} \sum_{k=0}^{n-1} F(\Phi(k)), \quad n \geq 1.$$

The chain $\{\Phi(k) : k \geq 0\}$ takes values in a general state space X , with transition kernel P , and it is assumed that the Lyapunov drift condition holds: $PV \leq V - W + b\mathbb{1}_C$ where $V: \mathsf{X} \rightarrow (0, \infty)$, $W: \mathsf{X} \rightarrow [1, \infty)$, the set C is small, and W dominates F . Under these assumptions, the following conclusions are obtained:

- (i) It is known that this drift condition is equivalent to the existence of a unique invariant distribution π satisfying $\pi(W) < \infty$, and the Law of Large Numbers holds for any function F dominated by W :

$$\phi_n \rightarrow \phi := \pi(F), \quad a.s., \quad n \rightarrow \infty.$$

- (ii) The lower error probability defined by $\mathbb{P}\{\phi_n \leq c\}$, for $c < \phi$, $n \geq 1$, satisfies a large deviation limit theorem when the function F satisfies a monotonicity condition. Under additional minor conditions an exact large deviations expansion is obtained.
- (iii) If W is near-monotone then control-variates are constructed based on the Lyapunov function V , providing a *pair of estimators* that together satisfy non-trivial large asymptotics for the lower *and* upper error probabilities.

In an application to simulation of queues it is shown that exact large deviation asymptotics are possible even when the estimator does not satisfy a Central Limit Theorem.

Keywords: Markov chains; Monte Carlo; control variate; large deviations.

2000 AMS Subject Classification: Primary: 60F10 65C05 37A30

Secondary: 60J22 00A72

*This paper is based upon work supported by the National Science Foundation under Award Nos. ECS 02 17836 and DMI 00 85165.

[†]Department of Electrical and Computer Engineering and the Coordinated Sciences Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, U.S.A. (meyn@uiuc.edu).

1 Introduction

This paper explores extensions of the control-variate method to obtain confidence bounds in simulation of a function of a Markov chain $\Phi = \{\Phi(0), \Phi(1), \dots\}$. It is assumed that Φ evolves on a general state space X , equipped with a countably generated sigma-field \mathcal{B} . The statistics of Φ are determined by its initial distribution, and the transition kernel P defined by,

$$P(x, A) := \mathbb{P}\{\Phi(1) \in A \mid \Phi(0) = x\}, \quad x \in \mathsf{X}, A \in \mathcal{B}.$$

Let $\{L_n : n \geq 1\}$ denote the sequence of empirical measures induced by Φ on $(\mathsf{X}, \mathcal{B})$,

$$L_n := \frac{1}{n} \sum_{k=0}^{n-1} \delta_{\Phi(k)} \quad n \geq 1. \quad (1)$$

It is assumed that Φ is positive Harris recurrent, with unique invariant probability distribution denoted π . Equivalently, for each bounded measurable function $F : \mathsf{X} \rightarrow \mathbb{R}$, and each initial condition, the Law of Large Numbers holds:

$$L_n(F) \rightarrow \phi := \pi(F) = \int F(x) \pi(dx), \quad n \rightarrow \infty, \text{ a.s.}$$

See [17, 14] or [31, Th. 17.1.7]. For each measurable function $F : \mathsf{X} \rightarrow \mathbb{R}$ satisfying $\pi(|F|) < \infty$, the sequence $\{L_n(F) : n \geq 1\}$ is interpreted as Monte-Carlo estimates of the steady-state mean of F .

While consistent for each initial condition even when F is not bounded, finer assumptions are required to obtain *confidence bounds*, i.e., bounds on $\mathbb{P}\{|L_n(F) - \phi| \geq a\}$ for a given $a > 0$. Such bounds are typically based on one of the following limit theorems:

The Central Limit Theorem, or CLT: For some $\sigma \geq 0$ and each initial condition,

$$\sqrt{n}[L_n(F) - \phi] \xrightarrow{w} \sigma X, \quad n \rightarrow \infty, \quad (2)$$

where X is a standard normal random variable, and the convergence is in distribution.

The Large Deviations Principle, or LDP: For a convex function $I : \mathbb{R} \rightarrow \mathbb{R}_+$, and any non-empty open interval $(c_0, c_1) \subset \mathbb{R}$,

$$\begin{aligned} & \lim_{n \rightarrow \infty} n^{-1} \log \mathbb{P}\{L_n(F) - \phi \in (c_0, c_1)\} \\ &= \lim_{n \rightarrow \infty} n^{-1} \log \mathbb{P}\{L_n(F) - \phi \in [c_0, c_1]\} = - \left(\min_{c \in [c_0, c_1]} I(c) \right). \end{aligned} \quad (3)$$

Sufficient as well as necessary conditions for the CLT are presented in [36, 15, 31, 17].

Following the work of Donsker and Varadhan [10, 39, 38] there is a significant literature on large deviations properties of Markov chains under a range of assumptions, and with a range of possible conclusions. In particular, precise limit theory under assumptions ranging from geometric ergodicity to Doeblin's condition are contained in [8, 9, 3, 26, 25]. The

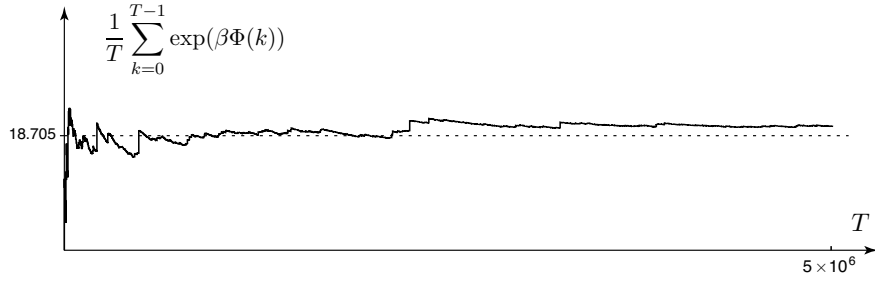


Figure 1: Monte-Carlo estimates of $\phi := \pi(F)$ with $F(x) = e^{0.1x}$ for $x \in \mathbb{Z}_+$. The stochastic process Φ is an M/M/1 queue initialized at zero, with load $\rho = 0.9$. After a transient period, the estimates are consistently larger than the steady-state mean of $\phi = (1 - \rho e^\beta)^{-1}(1 - \rho)$.

papers [34, 35, 6, 22, 11, 7] develop LDP *bounds* based on significantly milder assumptions. Explicit bounds on the rate function are obtained for uniformly ergodic chains in [16].

Recall that P acts on functions $h : \mathsf{X} \rightarrow \mathbb{R}$ and probability measures μ on \mathcal{B} , via

$$Ph(\cdot) = \int_{\mathsf{X}} P(\cdot, dy)h(y) \quad \text{and} \quad \mu P(\cdot) = \int_{\mathsf{X}} \mu(dx)P(x, \cdot). \quad (4)$$

The following bilinear and quadratic forms are defined for measurable functions $F, G : \mathsf{X} \rightarrow \mathbb{R}$,

$$\langle\langle F, G \rangle\rangle := P(FG) - (PF)(PG), \quad \mathcal{Q}(F) := P(F^2) - (PF)^2.$$

Under appropriate regularity assumptions, it is known that $\sigma^2(F) := \pi(\mathcal{Q}(ZF))$ is equal to the asymptotic variance given in (2), where Z is the *fundamental kernel* (see [31, Th. 17.5.3], and Proposition 2.2 below.)

The purpose of the control-variate method is to reduce the variance of the standard estimator defined by

$$\phi_n := L_n(F), \quad n \geq 1. \quad (5)$$

To describe this technique we fix a function $H : \mathsf{X} \rightarrow \mathbb{R}$ satisfying $\pi(H) = 0$, and set $F_\theta = F - \theta H$ so that $L_n(F_\theta)$ is an asymptotically unbiased estimator of ϕ for each $\theta \in \mathbb{R}$.

The asymptotic variance of the controlled estimator is given by

$$\sigma^2(F_\theta) = \pi(\mathcal{Q}(ZF_\theta)) = \pi(\langle\langle ZF, ZF \rangle\rangle - 2\theta\langle\langle ZF, ZH \rangle\rangle + \theta^2\langle\langle ZH, ZH \rangle\rangle).$$

Minimizing over $\theta \in \mathbb{R}$ gives the estimator with minimal asymptotic variance,

$$\theta^* = \frac{\pi(\langle\langle ZF, ZH \rangle\rangle)}{\pi(\langle\langle ZH, ZH \rangle\rangle)}.$$

See [29, 33, 13, 12] for more details and background on the general control-variate method.

An approach considered in [18, 20] is to consider functions of the form $H = J - PJ$, and choose J so that it approximates the solution \widehat{F} to *Poisson's equation*,

$$P\widehat{F} = \widehat{F} - F + \phi. \quad (6)$$

The idea is that if $J = \widehat{F}$, then the resulting controlled estimator with $\theta = 1$ has *zero* asymptotic variance.

This approach has been successfully applied in queueing models by taking J equal to an associated *fluid value function*. The approach is provably effective in simple models [18], and numerical examples show dramatic variance reduction for more complex networks [19]. Some theory to help explain the results of [19] is developed in [30] based on large-deviation limit results contained in [26].

For the sake of illustration consider the reflected random walk on \mathbb{R}_+ , defined by the recursion

$$\Phi(k+1) = [\Phi(k) + D(k+1)]_+, \quad k \geq 0, \quad (7)$$

with $[x]_+ = \max(x, 0)$ for $x \in \mathbb{R}$, and \mathbf{D} i.i.d.. We consider first the special case in which \mathbf{D} has common marginal distribution,

$$D(k) = \begin{cases} 1 & \text{with probability } \alpha; \\ -1 & \text{with probability } 1 - \alpha. \end{cases}$$

In this case Φ is a discrete-time model of the M/M/1 queue, and the state space is then restricted to $\mathbf{X} = \mathbb{Z}_+$. It is assumed that $\alpha \in (0, \frac{1}{2})$ so that Φ is a positive recurrent Markov chain on \mathbf{X} .

The invariant distribution π is geometric, so there is little motivation to simulate. However, ignoring this issue momentarily, suppose we wish to estimate using simulation the steady-state mean of $F(x) = e^{\beta x}$ for a given $\beta > 0$.

Shown in Figure 1 are Monte-Carlo estimates of the steady-state mean,

$$\phi := \sum \pi(i)F(i) = (1 - \rho) \sum \rho^i e^{\beta i},$$

where $\rho := \alpha/(1 - \alpha)$. We take $\beta = 0.1$, and $\alpha = 9/19$, so that $\rho = 9/10$, and $\phi = (1 - \rho e^\beta)^{-1}(1 - \rho) \approx 18.705$. The Markov chain Φ was initialized at zero, $\Phi(0) = 0$. The runlength in this simulation extended to $T = 5 \times 10^6$, yet the estimates are significantly larger than the steady state mean over much of the run. The following proposition provides some explanation. A proof is provided in Section 3.2. We note that the existence of non-trivial confidence bounds depends upon structure of the sublevel sets of F , defined by $C_F(r) := \{x : F(x) \leq r\}$ for $r \geq 1$. This structure holds here since the sublevel sets are finite for each r .

Proposition 1.1. *For the M/M/1 queue with $\rho = \alpha/(1 - \alpha) < 1$ and the function $F(x) = e^{\beta x}$, $x \in \mathbb{Z}_+$, with $1 < e^\beta < \rho^{-1}$, the steady-state mean is given by $\phi = (1 - \rho e^\beta)^{-1}(1 - \rho)$. The Monte-Carlo estimates of ϕ are consistent, and there exists $\bar{c} < \phi$ such that the LDP (3) holds for any open set $O \subset (\bar{c}, \infty)$, and each initial condition $\Phi(0) = x \in \mathbb{Z}_+$. The convex rate function $I: [\bar{c}, \infty) \rightarrow \mathbb{R}_+$ is strictly positive on $[\bar{c}, \phi)$, and identically zero on $[\phi, \infty)$. Consequently, we have for each initial condition $x \in \mathbb{Z}_+$,*

$$\begin{aligned} \lim_{n \rightarrow \infty} \log \left(\mathbb{P}_x \left\{ L_n(F) \leq c \right\} \right) &= -I(c) < 0, & c \in (\bar{c}, \phi), \\ \lim_{n \rightarrow \infty} \log \left(\mathbb{P}_x \left\{ L_n(F) \geq c \right\} \right) &= 0, & c \geq \phi. \end{aligned}$$

□

In this example the chain is geometrically ergodic, so an LDP bound might not be surprising [26]. In Section 3.2 we give a similar example in which analogous conclusions hold, yet $\{\phi_n\}$ does not even satisfy the CLT. Moreover, in this example we construct a control-variate to obtain a *pair* of estimators giving upper *and* lower confidence bounds.

The remainder of the paper is organized as follows. Section 2 contains a statement and proof of the most important conclusion in this paper, Theorem 2.1, which establishes the LDP for general class of functions on X . Some ergodic theory is reviewed in Section 2.1, and Section 2.2 contains a survey of spectral theory for Markov chains, following [32, 2, 26, 25]. A new criterion for the existence of a spectral gap is presented in Section 2.3, which is the main ingredient in the proof of Theorem 2.1.

Applications of Theorem 2.1 to the construction and analysis of control-variates are contained in Section 3.1. The simulation algorithm proposed in Section 3.1 is shown to satisfy exact upper and lower LDP bounds. This result is illustrated in Section 3.2 using the reflected random walk (7). Conclusions are contained in Section 4.

2 One-Sided Large Deviation Asymptotics

Throughout the paper we assume that Φ is positive Harris recurrent and aperiodic. Equivalently, there is a unique invariant probability distribution π on \mathcal{B} such that, for any $A \in \mathcal{B}$ satisfying $\pi(A) > 0$, and any initial condition x ,

$$\lim_{k \rightarrow \infty} \|P^k(x, \cdot) - \pi(\cdot)\| = 0, \quad x \in \mathsf{X},$$

where $\|\cdot\|$ denotes the total-variation norm [31, Th. 13.0.1]. We denote by \mathcal{B}^+ the set of $A \in \mathcal{B}$ satisfying $\pi(A) > 0$. We write $f \in \mathcal{B}^+$ if $f: \mathsf{X} \rightarrow \mathbb{R}_+$ is a measurable function with $\int f d\pi > 0$.

A measurable function $s: \mathsf{X} \rightarrow \mathbb{R}_+$ and a probability measure ν on \mathcal{B} are called *small* if for some $n \geq 1$ we have,

$$P^n(x, A) \geq s(x)\nu(A), \quad x \in \mathsf{X}, A \in \mathcal{B}. \quad (8)$$

A set C is called *small* if $s = \epsilon \mathbb{1}_C$ is a small function for some positive ϵ .

Under the condition that Φ is positive Harris recurrent it follows that there exists an extended-real valued $V: \mathsf{X} \rightarrow (0, \infty]$ that is finite a.e. $[\pi]$ and satisfies the following drift condition,

$$\left. \begin{array}{l} \text{For a function } W: \mathsf{X} \rightarrow [1, \infty), \text{ a small set } C \subset \mathsf{X}, \text{ and a constant } b < \infty, \\ PV \leq V - W + b\mathbb{1}_C, \quad \text{on } C_V := \{x : V(x) < \infty\}. \end{array} \right\} \quad (\mathbf{V3})$$

The set C_V is absorbing, so that the chain can be restricted to this set - see [31, Prop. 4.2.3]. In fact, given any measurable function $F: \mathsf{X} \rightarrow \mathbb{R}$ satisfying $\pi(|F|) < \infty$ we can construct a solution to (V3) with $W = 1 + |F|$ by applying [31, Th. 14.2.3].

We list below some other definitions for a given measurable function $F: \mathsf{X} \rightarrow \mathbb{R}$: The function is called,

Degenerate if there is a measurable function $H: \mathsf{X} \rightarrow \mathbb{R}$ such that when $\Phi(0) \sim \pi$ then $F(\Phi(k+1)) - F(\Phi(k)) = H(\Phi(k))$ a.s. for $k \geq 0$. Under appropriate bounds on H this implies that the asymptotic variance of ϕ_n is equal to zero. A converse is provided in [25, Lemma 4.12], based on [26, Prop. 2.4].

Lattice if there are $h > 0$ and $0 \leq d < h$, such that

$$\frac{F(x) - d}{h} \quad \text{is an integer,} \quad x \in \mathsf{X}. \quad (9)$$

If there exists a lattice function F_ℓ such that $F - F_\ell$ is degenerate, then F is called *almost-lattice*. Otherwise, F is called *strongly non-lattice*.

Near-monotone if $\inf_{x \in \mathsf{X}} F(x) > -\infty$, and the sublevel set $C_F(r) := \{x \in \mathsf{X} : F(x) \leq r\}$ is small or empty for each $r < \|F_+\|_\infty$, where $F_+ = \max(F, 0)$ and $\|\cdot\|_\infty$ denotes the supremum norm.

Large-deviation bounds are obtained in [2] for countable state-space chains under the assumption that F is near-monotone, and $\Lambda(F) < \|F\|_\infty$. These assumptions are far stronger than geometric ergodicity when F is unbounded. The results of [2] are strengthened and generalized to general state space chains and processes in [26, 25, 27].

By restricting the range of c in (3) we can relax the geometric ergodicity assumption. The proof of Theorem 2.1 is included at the end of this section.

Theorem 2.1. *Suppose that (V3) holds with V everywhere finite, and that $F \in L_\infty^W$ is a non-degenerate function satisfying $\pi(F) = 0$. Suppose moreover that the sublevel set $C_F(r)$ satisfies for some $r > \phi = 0$, $r_0 < \infty$,*

$$C_F(r) \subset C_V(r_0). \quad (10)$$

Then, there exists $\bar{c}_0 < \phi$ and a smooth convex function $I: (\bar{c}_0, \phi) \rightarrow (0, \infty)$ such that

- (i) *The LDP (3) holds for each initial condition $x \in \mathsf{X}$ and each $c \in (\bar{c}_0, \phi)$.*
- (ii) *If F is strongly non-lattice then for each $c \in (\bar{c}_0, \phi)$, there exists a bounded function $g_c: \mathsf{X} \rightarrow (0, \infty)$, such that for each initial condition $x \in \mathsf{X}$,*

$$\mathbb{P}_x\{L_n(F) \leq c\} \sim \frac{g_c(x)}{\sqrt{2\pi n}} e^{-nI(c)}, \quad n \rightarrow \infty. \quad (11)$$

□

The LDP asymptotics described in Theorem 2.1 are based on the spectral theory of a positive semigroup obtained from the function F to be simulated. The definitions presented in Section 2.2 are taken from [36, 31, 32, 2, 26, 25].

We first review some ergodic theory for Markov chains.

2.1 Ergodic theorems

For a given function $W: \mathsf{X} \rightarrow [1, \infty)$ the weighted L_∞ -norm is defined as

$$\|h\|_W := \sup_x |h(x)|/W(x),$$

and L_∞^W denotes the set of all measurable functions $h: \mathsf{X} \rightarrow \mathbb{R}$ for which this norm is finite (see [23, 24, 21, 31, 26, 25].) Note that the supremum norm $\|\cdot\|_\infty$ is precisely $\|\cdot\|_W$ with $W \equiv 1$.

Two functions $W, W': \mathsf{X} \rightarrow [1, \infty)$ are called *equivalent* if they generate the same function space, that is

$$W' \in L_\infty^W \text{ and } W \in L_\infty^{W'}.$$

The set of finite measures on \mathcal{B} is denoted \mathcal{M} ; the set $\mathcal{M}_1 \subset \mathcal{M}$ denotes probability measures on \mathcal{B} ; $\mathcal{M}^W \subset \mathcal{M}_1$ denotes measures satisfying $\mu(W) < \infty$; and $\mathcal{M}_1^W = \mathcal{M}^W \cap \mathcal{M}_1$.

The following ergodic properties for the chain are used in the analysis of simulation algorithms:

Proposition 2.2. *Suppose that Φ is ψ -irreducible and aperiodic and (V3) holds with V everywhere finite. Then,*

- (i) *The chain is positive Harris recurrent with $\pi(W) < \infty$, and we have*

$$\lim_{k \rightarrow \infty} \|P^k(x, \cdot) - \pi(\cdot)\|_W = 0, \quad x \in \mathsf{X}.$$

Moreover, the fundamental kernel $Z: L_\infty^W \rightarrow L_\infty^V$ exists as a bounded linear operator.

- (ii) *If $\pi(V) < \infty$ then the chain is Φ is called W -regular of degree two. In this case the fundamental kernel can be expressed as the sum,*

$$ZF = [I - (P - 1 \otimes \pi)]^{-1}F = \sum_{k=0}^{\infty} [P^k F - \phi], \quad F \in L_\infty^W.$$

The sum converges in the induced operator norm from L_∞^W to L_∞^V .

- (iii) *If $\pi(WV) < \infty$, then the CLT (2) holds for each $F \in L_\infty^W$.*

- (iv) *If $W \geq \varepsilon_0 V$ for some $\varepsilon_0 > 0$ then we say that (V4) holds. In this case Φ is V -uniformly ergodic: for some constants $b_0 < \infty$, $r_0 > 1$,*

$$\sum_{k=0}^{\infty} r_0^k \|P^k(x, \cdot) - \pi(\cdot)\|_V \leq b_0(V(x) + 1), \quad x \in \mathsf{X}.$$

Proof. The convergence result in Parts (i) and (ii) are contained in the f -Norm Ergodic Theorem (Theorem 14.0.1) of [31]. The interpretation of the sum as a version of the fundamental kernel is contained in [31, Th. 17.4.2].

Part (iii) follows from [31, Th. 17.4.4], and (iv) is contained in [31, Th. 16.0.1]. \square

2.2 Positive semigroups

Consider now a positive kernel \widehat{P} on $\mathsf{X} \times \mathcal{B}$. It is assumed that the semigroup $\{\widehat{P}^k : k \geq 0\}$ is ψ -irreducible,

$$\sum_{k=0}^{\infty} \widehat{P}^k(x, A) > 0, \quad x \in \mathsf{X}, \quad A \in \mathcal{B}^+,$$

and also *aperiodic*,

$$\widehat{P}^k(x, A) > 0, \quad \text{for all } x \in \mathsf{X}, \quad A \in \mathcal{B}^+, \quad \text{and all } k = k(x, A) \geq 1 \text{ sufficiently large.}$$

For a ψ -irreducible kernel there exists a function $s \in \mathcal{B}^+$, a probability measure ν on \mathcal{B} , and $n \geq 1$ such that $\widehat{P}^n \geq s \otimes \nu$. The function s and measure ν are called \widehat{P} -small, generalizing the definition for P .

Based on a given function $h: \mathsf{X} \rightarrow (0, \infty)$ we consider in Section 2.3 below the two positive kernels,

The *scaled kernel*: $P_h := I_h P$, or equivalently,

$$P_h(x, A) := h(x)P(x, A), \quad x \in \mathsf{X}, \quad A \in \mathcal{B}.$$

The *twisted kernel*: $\check{P}_h := I_{Ph}^{-1} P I_h$, or equivalently,

$$\check{P}_h(x, A) = \frac{\int_A P(x, dy) h(y)}{Ph(x)}, \quad x \in \mathsf{X}, \quad A \in \mathcal{B}. \quad (12)$$

The twisted kernel is *probabilistic*, so that $\check{P}_h(x, \mathsf{X}) = 1$ for all x , provided $Ph(x) < \infty$ for all $x \in \mathsf{X}$.

For any ψ -irreducible and aperiodic semigroup, the *generalized principal eigenvalue* (g.p.e.) is defined as $\lambda = e^\Lambda$, where $\Lambda \in [-\infty, \infty]$ is the limit,

$$\Lambda := \lim_{k \rightarrow \infty} k^{-1} \log(\nu \widehat{P}^k s) \quad (13)$$

The limit is independent of the particular small function $s \in \mathcal{B}^+$ and small measure ν chosen. If Λ is finite, then there is an associated eigenfunction $h: \mathsf{X} \rightarrow (0, \infty]$ satisfying $h(x) < \infty$ a.e. $[\psi]$, and

$$\widehat{P}h \leq \lambda h.$$

This is an equality provided \widehat{P} is λ -recurrent,

$$\lim_{k \rightarrow \infty} \lambda^{-k} \nu \widehat{P}^k s = \infty.$$

See [36, 7, 26] for further discussion.

For a given weighting function $v: \mathsf{X} \rightarrow [1, \infty)$, the induced operator norm of \widehat{P} is

$$\|\widehat{P}\|_v := \sup \left\{ \frac{\|\widehat{P}h\|_v}{\|h\|_v} : h \in L_\infty^v, \|h\|_v \neq 0 \right\}. \quad (14)$$

The *spectrum* $\mathcal{S}(\widehat{P}) \subset \mathbb{C}$ of \widehat{P} is the set of $z \in \mathbb{C}$ such that the inverse $[Iz - \widehat{P}]^{-1}$ does not exist as a bounded linear operator on L_∞^v .

The *spectral radius* of the semigroup $\{\widehat{P}^t\}$ is expressed $\xi = e^\Xi$, where

$$\Xi := \lim_{k \rightarrow \infty} k^{-1} \log(\|\widehat{P}^k\|_v) \quad (15)$$

We say that \widehat{P} is *v-uniform* if the spectral radius ξ is finite, and there exists $h \in L_\infty^v$, $\mu \in \mathcal{M}_1^v$, such that

$$\sup_{|z| \geq \xi} \|[Iz - (\widehat{P} - h \otimes \mu)]^{-1}\|_v < \infty.$$

When $v \equiv 1$ we drop the qualification and simply say that \widehat{P} is *uniform*.

If the kernel is *v-uniform* then it admits a spectral gap, and the generalized principle eigenvalue coincides with ξ . Moreover, the eigenfunction h satisfies $h \in L_\infty^v$, and the eigenfunction equation $\widehat{P}h = \lambda h$ holds [25, Prop. 2.9].

2.3 Multiplicative mean ergodic theorem

The multiplicative mean ergodic theorem contained in Theorem 2.3 is the basis of LDP asymptotics for the partial sums [2, 26, 25].

The proof of Theorem 2.3 is identical to the proof of Theorem 3.4 in [25]. The idea of the proof of (17) is as follows: The twisted kernel \check{P}_h is \check{v} -uniform, where $h = \check{f}$ is an eigenfunction and $\check{v} := v/h$, since the twisted kernel is simply a scaling and similarity transformation of P_f . Since each of the twisted kernels is probabilistic, this implies that \check{P}_h is the transition kernel for a \check{v} -uniformly ergodic Markov chain (see [26, Cor. 4.7], or [25, Prop. 2.11] for finer results.) The multiplicative mean-ergodic theorem is a consequence of this mean-ergodic theorem for the twisted chain.

An explicit formula for the eigenfunction is given in (25) below.

Theorem 2.3. *Suppose that $F: \mathbf{X} \rightarrow \mathbb{R}$ is measurable; its g.p.e. λ is finite; and that P_f is v -uniform, with $f = e^F$. Then, there exists a measure $\check{\mu} \in \mathcal{M}_1^v$ and a function $\check{f} \in L_\infty^v$ satisfying the eigenfunction equations $\widehat{P}\check{f} = \lambda\check{f}$, $\check{\mu}\widehat{P} = \lambda\check{\mu}$, with normalization,*

$$\check{\mu}(\check{f}) = \check{\mu}(\mathbf{X}) = 1, \quad (16)$$

and these are the unique solutions. Moreover, the following multiplicative mean ergodic theorem holds: For some $b_0 > 0$, $b_1 < \infty$, and all $x \in \mathbf{X}$, $n \geq 1$,

$$\left| \mathbf{E}_x \left[\exp \left(\sum_{k=0}^{n-1} F(\Phi(k)) - n\Lambda \right) \right] - \check{f}(x) \right| \leq b_1 e^{-b_0 n} v(x). \quad (17)$$

□

In the series of results that follow we present sufficient conditions for uniformity of a scaled kernel. It will be convenient to consider a family of kernels $\{P_a : a \in (0, 1]\}$ where

$P_a = P_{f_a}$ is the scaled kernel defined with $f_a = e^{aF}$, and $F: X \rightarrow \mathbb{R}$ a given measurable function. A family of resolvent kernels is defined by,

$$R_a := \sum_{k=0}^{\infty} 2^{-k-1} P_a^k, \quad a \in (0, 1], \quad (18)$$

and R denotes the kernel obtained when $a = 0$ so that $f_a \equiv 1$. It can be shown that v -uniformity of R_a is equivalent to v -uniformity of P_a when P is aperiodic and $\lambda_a < 2$, and in this case λ_a is the g.p.e. for P_a if and only if $\gamma_a = (2 - \lambda_a)^{-1}$ is the g.p.e. for R_a .

We assume throughout that the function F is normalized so that $\pi(F) = 0$. It then follows from the definition (13), Jensen's inequality, and the mean ergodic theorem that $\lambda_a, \gamma_a \in [1, \infty]$ for each $a \in \mathbb{R}$.

Under the assumptions of Theorem 2.4 the kernel R_a is a bounded linear operator on $L_{\infty}^{v_a}$ for $a \in [0, 1]$, where $v_a = e^{aV}$. The first bound in (19) is analogous to Condition (V4) of [31]. These two bounds are equivalent to geometric ergodicity for the Markov chain in the special case $F \equiv 0$. For general F this is not true, as we shall see in Theorem 2.6.

Theorem 2.4. *Suppose that $F: X \rightarrow \mathbb{R}$ is a given measurable function satisfying $\pi(|F|) < \infty$ and $\pi(F) = 0$. Suppose that there exists $\bar{\lambda}_1 < 1$, a function $v: X \rightarrow [1, \infty]$ that is not everywhere infinite, a function $s: X \rightarrow \mathbb{R}_+$, a probability distribution ν on \mathcal{B} , and a constant $b < \infty$ satisfying the bounds,*

$$P_f v \leq \bar{\lambda}_1 v + bs, \quad (19)$$

and $R_a \geq s \otimes \nu$ for all $0 \leq a \leq 1$.

Then, $v(x) < \infty$ a.e. $[\pi]$, and there exists $\bar{a} \in (0, 1)$ such that P_a is v_a -uniform for all $a \in (0, \bar{a})$.

Proof. To show that R_a is v_a -uniform we prove that $\|G_a\|_{v_a} < \infty$, where G_a denotes the potential kernel,

$$G_a = \sum_{k=0}^{\infty} \gamma_a^{-k-1} [R_a - s \otimes \nu]^k, \quad (20)$$

and $\gamma_a = (2 - \lambda_a)^{-1}$ is the g.p.e. for R_a .

Jensen's inequality implies the following family of bounds,

$$P_a v_a \leq (\bar{\lambda}_1 v + bs)^a \leq \bar{\lambda}_a v_a + a \bar{\lambda}_{a-1} bs, \quad 0 \leq a \leq 1,$$

where $\bar{\lambda}_t := (\bar{\lambda}_1)^t$ for any $t \in \mathbb{R}$. Moreover, the resolvent equation holds: $P_a R_a = R_a P_a = 2R_a - I$. This combined with the bound on $P_a v_a$ gives,

$$2R_a v_a - v_a = R_a P_a v_a \leq R_a [\bar{\lambda}_a v_a + a \bar{\lambda}_{a-1} bs], \quad (21)$$

which on re-arranging terms implies the bound,

$$R_a v_a \leq \bar{\gamma}_a v_a + ab_a R_a s, \quad (22)$$

with $\bar{\gamma}_a = (2 - \bar{\lambda}_a)^{-1}$, and $b_a = b\bar{\lambda}_{a-1}\bar{\gamma}_a$. We evidently have $\bar{\gamma}_a < 1$ and $\bar{\lambda}_a < 1$, so that $b_a \leq b/\bar{\lambda}_1$ for $a \in [0, 1]$.

Define $v'_a = (1 + ab_2)v_a - ab_2s$ where $b_2 > b/\bar{\lambda}_1$ is fixed. This function is equivalent to v_a , and from (22),

$$R_a v'_a \leq (1 + ab_2)[\bar{\gamma}_a v_a + ab_a R_a s] - ab_2 R_a s.$$

For $a > 0$ sufficiently small we have $b_2 \geq (1 + ab_2)b/\bar{\lambda}_1 \geq (1 + ab_2)b_a$. Consequently, for such a ,

$$R_a v'_a \leq (1 + ab_2)\bar{\gamma}_a v_a = \bar{\gamma}_a v'_a + ab_2 \bar{\gamma}_a s,$$

and on subtracting the function $\nu(v'_a)s$ from each side this gives,

$$[R_a - s \otimes \nu]v'_a \leq \bar{\gamma}_a v'_a - (\nu(v'_a) - ab_2 \bar{\gamma}_a)s.$$

Decreasing a still further we can assume that $(\nu(v'_a) - ab_2 \bar{\gamma}_a) \geq 0$. We conclude that there exists $\bar{a} > 0$ such that with $\delta_a := 1 - \bar{\gamma}_a$,

$$[R_a - s \otimes \nu]v'_a \leq v'_a - \delta_a v'_a, \quad a \in (0, \bar{a}].$$

Iterating this bound gives,

$$[R_a - s \otimes \nu]^n v'_a \leq v'_a - \delta_a \sum_{k=0}^{n-1} [R_a - s \otimes \nu]^k v'_a, \quad n \geq 1,$$

and hence, $\sum_{k=0}^{\infty} [R_a - s \otimes \nu]^k v'_a \leq v'_a$, which implies the final bound,

$$\| [Iz - (R_a - s \otimes \nu)]^{-1} \|_{v'_a} \leq \delta_a^{-1}, \quad a \in (0, \bar{a}], \quad |z| \geq 1.$$

This completes the proof that $\|G_a\|_{v_a} < \infty$ since $\gamma_a \geq 1$, and v_a is equivalent to v'_a for $a \in (0, \bar{a}]$. \square

The following result provides a simple criterion that guarantees the existence of s, ν satisfying the minorization condition in (19).

Proposition 2.5. *Suppose that (V3) holds, and that $F \in L_{\infty}^W$. Then, for each $r \geq 1$, $a \in \mathbb{R}$, the set $C = C_V(r) := \{x : V(x) \leq r\}$ is small for the positive kernel P_a . Moreover, we have the following uniform bounds: For each $a_0 > 0$, $r \geq 1$, there exists $\varepsilon_0 > 0$, $n_0 \geq 1$, and a probability distribution $\nu_0 \in \mathcal{M}_1$ such that,*

$$P_a^{n_0}(x, A) \geq \varepsilon_0 \nu_0(A), \quad x \in C, \quad a \in [-a_0, a_0].$$

Proof. Fix $a_0 > 0$, $r \geq 1$, set $F_0(x) = a_0 \|F\|_W W(x)$, $x \in \mathsf{X}$, and define $\hat{P} := I_{e^{-F_0}} P$. This is simply the scaled kernel P_h with $h = e^{-F_0}$. A minorization condition obtained for the kernel \hat{P} will imply the desired uniform bounds since $P_a \geq \hat{P}$ for $a \in [-a_0, a_0]$.

For any $A \in \mathcal{B}$ and $x \in \mathsf{X}$ we have by Jensen's inequality,

$$\begin{aligned} \hat{P}^n(x, A) &= \left\{ P^n(x, A)^{-1} \mathbf{E}_x \left[\exp \left(- \sum_{k=0}^{n-1} F_0(\Phi(k)) \right) \mathbb{I}\{\Phi(n) \in A\} \right] \right\} P^n(x, A) \\ &\geq \exp \left\{ P^n(x, A)^{-1} \mathbf{E}_x \left[\left(- \sum_{k=0}^{n-1} F_0(\Phi(k)) \right) \mathbb{I}\{\Phi(n) \in A\} \right] \right\} P^n(x, A). \end{aligned} \tag{23}$$

Under (V3) the following bound holds, $0 \leq P^n V \leq V + nb - \sum_{k=0}^{n-1} P^k W$, so that

$$\mathbb{E}_x \left[\sum_{k=0}^{n-1} F_0(\Phi(k)) \right] \leq a_0 \|F\|_W [V(x) + nb], \quad x \in \mathbf{X}, n \geq 1.$$

Consequently, from (23),

$$\widehat{P}^n(x, A) \geq \exp \left\{ -P^n(x, A)^{-1} a_0 \|F\|_W [V(x) + nb] \right\} P^n(x, A), \quad x \in \mathbf{X}, n \geq 1. \quad (24)$$

This shows that the semigroup $\{\widehat{P}^n : n \geq 1\}$ is π -irreducible, in the sense that $\sum_n \widehat{P}^n(x, A) > 0$ for each x whenever $\pi(A) > 0$. Let $A \in \mathcal{B}^+$ be any fixed small set for \widehat{P} : there is a probability distribution ν_0 , $\varepsilon > 0$, and an integer $m \geq 1$ such that

$$\widehat{P}^m(x, B) \geq \varepsilon \nu_0(B), \quad x \in A, B \in \mathcal{B}.$$

Choose $n \geq 1$, $\delta > 0$ that $P^n(x, A) \geq \delta$ for $x \in C$. This is possible since the set C is W -regular, and hence small [31, Th. 14.2.3]. It follows from (24) that

$$\widehat{P}^n(x, A) \geq \delta_r := \exp \left\{ -\delta^{-1} a_0 \|F\|_W [r + nb] \right\} \delta, \quad x \in C,$$

and hence,

$$\widehat{P}^{n+m}(x, B) \geq \delta_r \varepsilon \nu_0(B), \quad x \in C, B \in \mathcal{B}.$$

This completes the proof with $n_0 = m + n$ and $\varepsilon_0 = \delta_r \varepsilon$. \square

Up to now it appears that uniformity is a tremendously strong assumption on the scaled kernel P_f since the implications of uniformity are so strong. However, under the assumptions of Theorem 2.1 we can establish uniformity of P_a for a range of *non-positive* a , even though Φ is only positive Harris recurrent.

Recall that \widehat{P} is called *uniform* if it is v -uniform with $v \equiv 1$.

Theorem 2.6. *Suppose that (V3) holds with V everywhere finite, and suppose that the function $F \in L_\infty^W$ satisfies (10) for some $r > 0$ and $r_0 < \infty$, with $\phi = \pi(F) = 0$. Then, there exists $\bar{a} < 0$ such that*

- (i) P_a is uniform for each $a \in (\bar{a}, 0)$.
- (ii) The eigenfunctions $\{\check{f}_a : a \in (\bar{a}, 0)\} \subset L_\infty$, normalized so that $\nu(\check{f}_a) = 1$ for some small measure ν and each a , are uniformly bounded:

$$\sup_{\substack{\bar{a} < a < 0 \\ x \in \mathbf{X}}} \check{f}_a(x) < \infty.$$

- (iii) Define $\check{f}'_a := \frac{d}{da} \check{f}_a$ for $a \in (\bar{a}, 0)$, with $\{\check{f}_a\}$ normalized as in (ii). These functions are uniformly bounded in norm:

$$\sup_{\bar{a} < a < 0} \|\check{f}'_a\|_V < \infty.$$

(iv) Λ is convex and analytic on $(\bar{a}, 0)$, and $\lim_{a \uparrow 0} \frac{d}{da} \Lambda(a) = 0$.

Define the twisted kernel by $\check{P}_a := \check{P}_{\check{f}_a}$, where \check{f}_a is an eigenfunction exists for P_a . We have noted prior to Theorem 2.3 that \check{P}_a is the transition kernel for a \check{v}_a -uniformly ergodic Markov chain when P_a is v_a -uniform. As in [26, Prop. 4.9] one can verify that each of the functions

$$\widehat{F}_a = \frac{d}{da} \log(\check{f}_a) = \check{f}'_a / \check{f}_a, \quad a \in (\bar{a}, 0),$$

solves Poisson's equation for the corresponding twisted kernel,

$$\check{P}_a \widehat{F}_a = \widehat{F}_a - F + \phi(a)$$

where $\phi(a) = \frac{d}{da} \Lambda(a)$ is the steady-state mean of F for the twisted kernel. Poisson's equation for \check{P}_a is used to establish versions of Theorem 2.4 (iv) in the papers [26, 25]. This technique cannot be applied here since we do not know if Λ is bounded or smooth for positive a .

The proof of Theorem 2.1 is performed in the remainder of this subsection through a series of steps.

We see in Lemma 2.7 that Part (i) of Theorem 2.1 follows quickly from Theorem 2.4. Recall that $G_a = [\gamma_a I - (R_a - s \otimes \nu)]^{-1}$ is the potential kernel previously defined in (20).

Lemma 2.7. *If the assumptions of Theorem 2.6 hold then there exists $\bar{a}_0 < 0$ such that P_a is uniform for each $a \in (\bar{a}_0, 0)$. The unique eigenfunction $\check{f}_a \in L_\infty$ satisfying $\nu(\check{f}_a) = 1$ can be expressed,*

$$\check{f}_a = G_a s. \quad (25)$$

Proof. Set $G = -F$, $g = e^G$, and $g_a = e^{aG}$ for $a \in \mathbb{R}$. Note that uniformity of P_{g_a} is equivalent to uniformity of P_{-a} for any a .

Define $v \equiv 1$, $\delta = r - \phi$, and $b = \exp(\sup_{x \in C_F(r)} |G(x)|) < \infty$, so that the following bound holds:

$$P_g v = e^{-F} \leq e^{-\delta} v + b \mathbb{1}_{C_F(r)}.$$

Moreover, Proposition 2.5 implies that the minorization condition in (19) holds with F replaced by G in the definition of R_a . Uniformity of P_{g_a} for sufficiently small $a > 0$ thus follows from Theorem 2.4 and the fact that v is bounded.

The representation (25) follows from Proposition 2.8 of [25] (see also [36].) \square

The difficult part of the proof of Theorem 2.6 is to establish convergence of $\phi(a)$ to ϕ as $a \uparrow 0$. The proof is based on consideration of the following scaled kernel to bound P_a .

For a given small measure $\nu \in \mathcal{M}_1$, let $s = \varepsilon_v \mathbb{1}_{C_V(r_0)}$ with $\varepsilon_v > 0$ chosen so that $R \geq s \otimes \nu$. For a fixed $\varepsilon > 0$, define the scaled kernel,

$$\bar{P}(x, A) := \exp(\varepsilon_v \varepsilon \mathbb{1}_{C_V(r_0)}(x)) P(x, A), \quad x \in \mathbf{X}, A \in \mathcal{B},$$

and the resolvent and potential kernels,

$$\bar{R} := \sum_{k=0}^{\infty} 2^{-k-1} \bar{P}^k, \quad \bar{G} := \sum_{k=0}^{\infty} [\bar{R} - s \otimes \nu]^k.$$

Lemma 2.8. *Suppose that the assumptions of Theorem 2.6 hold. Then, there exists $\varepsilon > 0$ such that*

$$(i) \|\bar{G}s\|_\infty < \infty, \quad (ii) \|\bar{G}\bar{R}s\|_\infty < \infty, \quad (iii) \|\bar{G}W\|_V < \infty.$$

Proof. To see (i) we write,

$$\bar{P}1 = 1 + (e^{\varepsilon\varepsilon v} - 1)\mathbb{I}_{C_V(r_0)} = 1 + \varepsilon_v^{-1}(e^{\varepsilon\varepsilon v} - 1)s.$$

Let $h = 1 - \varepsilon_v^{-1}(e^{\varepsilon\varepsilon v} - 1)s$, and choose $\varepsilon > 0$ so this is strictly positive everywhere. From the resolvent equation as in (21) we have

$$\bar{R}1 = 1 + \varepsilon_v^{-1}(e^{\varepsilon\varepsilon v} - 1)\bar{R}s,$$

and hence

$$\bar{R}h = 1 = h + \varepsilon_v^{-1}(e^{\varepsilon\varepsilon v} - 1)s.$$

On subtracting $[s \otimes \nu]h$ from both sides we then obtain,

$$(\bar{R} - s \otimes \nu)h = h - \delta_h s,$$

where $\delta_h = \nu(h) - \varepsilon_v^{-1}(e^{\varepsilon\varepsilon v} - 1)$. By reducing $\varepsilon > 0$ we can assume that $\delta_h > 0$.

Exactly as in the proof of Theorem 2.4 we conclude that

$$\bar{G}s := \sum_{k=0}^{\infty} [\bar{R} - s \otimes \nu]^k s \leq \delta_h^{-1} h,$$

which establishes the uniform bound in (i).

Part (ii) follows from (i) and the identity,

$$\bar{G}\bar{R} = \bar{G}[\bar{R} - s \otimes \nu] + \bar{G}[s \otimes \nu] = \bar{G} - I + \bar{G}[s \otimes \nu], \quad (26)$$

so that $\bar{G}\bar{R}s \leq [\bar{G} + (\bar{G}s) \otimes \nu]s = (1 + \nu(s))\bar{G}s$.

To see (iii) we note that we can assume without loss of generality that the set C in (V3) is equal to the set $C_V(r_0)$ used here by applying [31, Th. 14.2.3]. Under this transformation we obtain,

$$\begin{aligned} \bar{P}V &\leq V - W + (e^{\varepsilon s} - 1)V + (e^{\varepsilon\varepsilon v} - 1)b\mathbb{I}_C \\ &\leq V - W + b_v s, \end{aligned}$$

where $b_v = \varepsilon_v^{-1}(e^{\varepsilon\varepsilon v} - 1)(b + r_0)$.

From the resolvent equation again this implies the bound,

$$\bar{R}V \leq V - \bar{R}W + b_v \bar{R}s,$$

and then through familiar arguments, $\bar{G}\bar{R}W \leq V + b_v \bar{G}\bar{R}s$. This bound combined with the identity (26) completes the proof of (iii). \square

With this value of $\varepsilon > 0$ fixed in the definition of \bar{R} , and given $\bar{a}_0 < 0$ from Lemma 2.7, we now identify the lower bound \bar{a} :

Lemma 2.9. *Suppose that the assumptions of Theorem 2.6 hold. Then, there exists $\bar{a} \in (\bar{a}_0, 0)$ such that for any $\bar{a} < a \leq 0$,*

$$\lambda_a^{-1}P_a \leq P_a \leq \bar{P} \quad \text{and} \quad \check{f}_a \leq \bar{f} := \bar{G}s,$$

where \check{f}_a is defined in (25).

Proof. The bound $\lambda_a^{-1}P_a \leq P_a$ holds since $\lambda_a \geq 1$.

To see that $P_a \leq \bar{P}$, rewrite this bound as $f_a(x) \leq \exp(\varepsilon_v \varepsilon \mathbb{I}_{C_V(r_0)}(x))$, or on taking logarithms,

$$aF(x) \leq \varepsilon_v \varepsilon \mathbb{I}_{C_V(r_0)}(x), \quad x \in \mathbf{X}. \quad (27)$$

Letting $b_0 = \sup_{x \in C_0(F)} |F(x)|$ with $C_0(F) = \{x \in \mathbf{X} : F(x) \leq 0\}$, and applying the bound (10) gives,

$$aF(x) \leq |a|b_0 \mathbb{I}_{C_0(F)} \leq |a|b_0 \mathbb{I}_{C_V(r_0)}, \quad a < 0.$$

This shows that (27) holds for $a \in [\bar{a}, 0)$ with,

$$\bar{a} := \max(\bar{a}_0, -\varepsilon_v \varepsilon b_0^{-1}).$$

The second bound $\check{f}_a \leq \bar{f}$ follows immediately from the first since,

$$\gamma_a^{-n-1}(R_a - s \otimes \nu)^n \leq (R_a - s \otimes \nu)^n \leq (\bar{R} - s \otimes \nu)^n, \quad n \geq 0.$$

□

The previous two lemmas lead to a proof of Theorem 2.6 (iii):

Lemma 2.10. *Under the assumptions of Theorem 2.6 the function \check{f}'_a can be expressed,*

$$\check{f}'_a = G_a[\gamma_a \lambda_a R_a I_F - \lambda'_a I] \check{f}_a, \quad \bar{a} < a < 0,$$

where \bar{a} is defined in Lemma 2.9. These functions satisfy the uniform bound,

$$\sup_{\bar{a} < a < 0} \|\check{f}'_a\|_V < \infty.$$

Proof. The eigenfunction normalized with $\nu(\check{f}_a) = 1$ can be expressed as (25), where the potential kernel G_a exists as a bounded linear operator from L_∞^W to L_∞^V by the previous two lemmas.

For any two values a_1, a_2 we then have,

$$\check{f}_{a_2} - \check{f}_{a_1} = G_{a_2}s - G_{a_1}s = G_{a_2}[(\gamma_{a_1} - \gamma_{a_2})I - (R_{a_1} - R_{a_2})]G_{a_1}s. \quad (28)$$

Lemma 2.9 and Lemma 2.8 imply that $\check{f}_{a_1} = G_{a_1}s$ is uniformly bounded. Convexity of Λ implies that $|\gamma_{a_1} - \gamma_{a_2}|(a_2 - a_1)^{-1}$ is uniformly bounded for $\bar{a} \leq a_2 < a_1 < 0$, and it may then be verified using the Mean Value Theorem that for some constant $b_0 < \infty$,

$$(a_2 - a_1)^{-1} \left(|\gamma_{a_1} - \gamma_{a_2}| + \|R_{a_1}1 - R_{a_2}1\|_W \right) \leq b_0, \quad \bar{a} \leq a_2 < a_1 < 0.$$

Applying Lemma 2.8 once more we conclude that

$$(a_2 - a_1)^{-1} \|\check{f}_{a_2} - \check{f}_{a_1}\|_V \leq b_1, \quad \bar{a} \leq a_2 < a_1 < 0,$$

where $b_1 := b_0 \|\bar{G}W\|_V < \infty$. These bounds justify considering the limit $a_2 \downarrow a_1$ in (28) to obtain both the desired expression for \check{f}'_a and the uniform bounds. □

Lemma 2.11. *Suppose that the assumptions of Theorem 2.6 hold. Then $\check{f}_a \rightarrow 1$ and $\lambda_a \rightarrow 1$ as $a \uparrow 0$.*

Proof. This follows from the uniform bound $G_a \leq \bar{R}$, and the formula for the limiting potential kernel,

$$G = \sum_{k=0}^{\infty} [R - s \otimes \nu]^k.$$

We have $\check{f}_a = G_a s \rightarrow Gs$ as $a \uparrow 0$, and it is known that $Gs \equiv 1$ since Φ is Harris recurrent [36, Th. 5.1]. \square

Proof of Theorem 2.6. Part (i) is given in Lemma 2.7; (ii) follows from Lemma 2.8 and Lemma 2.9; and (iii) is given in Lemma 2.10.

To see that Λ is smooth we argue as in [26, 25]: the g.p.e. γ_a for R_a is defined for $a \in (\bar{a}, 0)$ as the unique solution to

$$\nu[I\gamma - (R_a - s \otimes \nu)]^{-1}s = 1.$$

Hence smoothness follows from the Inverse Function Theorem - see Proposition 4.8 of [26].

We now show that $\phi(a) = \Lambda'(a) \rightarrow \phi$ as $a \uparrow 0$. Based on Lemma 2.10 we have,

$$h := \lim_{a \uparrow 0} \check{f}'_a = GX\check{f}_{0-},$$

with $X := (\gamma_{0-} - \lambda_{0-})RI_F - \lambda'_{0-}I$. Moreover, $\nu(\check{f}'_a) = 0$ for each a , and hence from the uniform bounds on \check{f}'_a combined with the Dominated Convergence Theorem we have $\nu(h) = 0$.

Lemma 2.11 implies that $\gamma_{0-} = \lambda_{0-} = 1$ so that $X = [RI_F - \lambda'_{0-}I]$, and also $\check{f}_{0-} = 1$ so that

$$0 = \nu(h) = \nu GX1 = \mu(RF - \lambda'_{0-})$$

where $\mu = \nu G$ is an un-normalized invariant measure. In particular $\mu R = \mu$, so that the expression above implies that $\mu(F) = \lambda'_{0-}\mu(X)$, or equivalently,

$$\phi := \pi(F) = \frac{\mu(F)}{\mu(X)} = \lambda'_{0-}.$$

\square

2.4 Proof of Theorem 2.1

Under the assumptions of Theorem 2.6 the function $\Lambda(a)$ is convex on $(\bar{a}, 0)$. Define the parameters,

$$\bar{c}_0 = \lim_{a \downarrow \bar{a}} \frac{d}{da} \Lambda(a); \quad \bar{c}_1 = \lim_{a \uparrow 0} \frac{d}{da} \Lambda(a).$$

Theorem 2.6 (iv) implies that $\bar{c}_1 = \phi$. By convexity we have $\bar{c}_0 \leq \bar{c}_1$, and this inequality is strict if F is non-degenerate since then

$$\sigma_a^2 := \frac{d^2}{da^2} \Lambda(a) > 0, \quad a \in (\bar{a}, 0). \tag{29}$$

Equation (29) follows from [26, Prop. 2.4] (see also [25, Lemma 4.12].)

For $c \in (\bar{c}_0, \bar{c}_1)$ the convex dual of Λ is expressed

$$I(c) = \max_{a \in (\bar{a}, 0)} [ca - \Lambda(a)] = ca^* - \Lambda(a^*)$$

where a^* is chosen so that $\frac{d}{da}\Lambda(a) = c$. The function I serves as an LDP rate function within this range.

Since Φ satisfies the assumptions of Theorem 2.4 it follows from Theorem 2.6 that $\bar{c}_1 = \phi$. The LDP (3) for $c \in (\bar{c}_0, \phi)$ then follows from the multiplicative ergodic theorem Theorem 2.3 and standard arguments [8].

The proof of the exact LDP (11) is identical to that of the corresponding results in [26, 25], where

$$g_c = \frac{1}{a^* \sigma_{a^*}} \check{f}_{a^*},$$

$a^* \in (\bar{a}, 0)$ is again chosen so that $\frac{d}{da}\Lambda(a) = c$, \check{f}_{a^*} is the eigenfunction satisfying the normalization (16), and $\sigma_{a^*}^2$ is defined in (29). \square

3 Application to control variates

In this section we show how the method of control variates can be used to construct a simulator that satisfies an LDP for the upper and lower tails even when the assumptions on F in Theorem 2.1 are violated.

We begin with a general application of Theorem 2.1.

3.1 Control variates based on a Lyapunov function

Suppose that (V3) holds, and consider the function $H := V - PV$. If $\pi(V) < \infty$ then invariance of π implies that $\pi(H) = 0$, and hence the function H can be used to construct a control-variate as described in the Introduction.

Recall that the assumption $\pi(V) < \infty$ means that the chain is W -regular of degree two. In the following result we prefer to avoid this restriction and simply assume directly that $\pi(|H|) < \infty$ and that $\pi(H) = 0$. Under this assumption, with $H := V - PV$, define the sequence,

$$\Delta_n := n^{-1} \sum_{k=0}^{n-1} H(\Phi(k)), \quad n \geq 0. \quad (30)$$

Positive Harris recurrence of Φ implies that $\Delta_n \rightarrow 0$ as $n \rightarrow \infty$ with probability one [31, Th. 17.0.1]. The control variate for simulation of a function F based on the control variate H and a given parameter $\theta \in \mathbb{R}$ is given by,

$$L_n(F_\theta) := L_n(F) - \theta L_n(H) = L_n(F) - \theta \Delta_n, \quad n \geq 1.$$

In Theorem 3.1 we fix a function $F \in L_\infty^W$ together with constants θ_-, θ_+ each strictly greater than $\|F\|_W$. Define,

$$F_- = F - \theta_- H, \quad F_+ = F + \theta_+ H,$$

and the pair of estimators,

$$\phi_n^- = L_n(F_-), \quad \phi_n^+ = L_n(F_+), \quad n \geq 1. \quad (31)$$

We denote by $\Lambda_-(a)$, $\Lambda_+(a)$ the logarithm of the generalized principal eigenvalue for each of the scaled kernels

$$P_a^- = I_{e^{aF_-}} P, \quad P_a^+ = I_{e^{aF_+}} P, \quad a \in \mathbb{R}.$$

The convex duals of the functions $\{\Lambda_-, \Lambda_+\}$ are denoted $\{I_-, I_+\}$.

Theorem 3.1. *Suppose that (V3) holds with $W: \mathsf{X} \rightarrow [1, \infty)$ near-monotone and $V: \mathsf{X} \rightarrow (0, \infty)$ everywhere finite, and suppose that $\pi(H) = 0$, where $H := V - PV$. Then, for a given function $F \in L_\infty^W$, there exists $\varepsilon_0 > 0$ such that*

(i) *The lower LDP limit holds using $\{\phi_n^+\}$:*

$$\lim_{n \rightarrow \infty} n^{-1} \log \mathbf{P}\{\phi_n^+ \leq c\} = -I_+(c), \quad c \in (\phi - \varepsilon_0, \phi).$$

(ii) *The upper LDP limit holds using $\{\phi_n^-\}$:*

$$\lim_{n \rightarrow \infty} n^{-1} \log \mathbf{P}\{\phi_n^- \geq c\} = -I_-(c), \quad c \in (\phi, \phi + \varepsilon_0).$$

(iii) *If in addition F is strongly non-lattice, then (i) and (ii) can be strengthened to the corresponding exact LDP limit analogous to (11).*

Parts (i) and (ii) of Theorem 3.1 combined imply that

$$\begin{aligned} \lim_{n \rightarrow \infty} n^{-1} \log \mathbf{P}_x \left\{ [\phi_n - \phi] \in [-\varepsilon - \theta_+ \Delta_n, \varepsilon + \theta_- \Delta_n]^c \right\} \\ = -\min(I_+(\phi - \varepsilon), I_-(\phi + \varepsilon)) < 0, \quad 0 < \varepsilon < \varepsilon_0. \end{aligned}$$

Proof of Theorem 3.1. By normalization we can assume without loss of generality that $\pi(F) = 0$.

Define $W' = H + b\mathbb{1}_C$ so that by definition of H we have $PV = V - W' + b\mathbb{1}_C$, and by (V3) we also have the lower bound,

$$W' = (V - PV) + b\mathbb{1}_C \geq (W - b\mathbb{1}_C) + b\mathbb{1}_C = W.$$

Both F and H belong to the function space $L_\infty^{W'}$.

We can write,

$$F_- = F - \theta_- H = F - \theta_- (W' - b\mathbb{1}_C),$$

which implies that $-F_-$ is near-monotone whenever $\theta_- > \|F\|_W$ since we have the explicit bound,

$$-F_- \geq (\theta_- - \|F\|_W)W - \theta_- b\mathbb{1}_C.$$

Similarly, F_+ is near-monotone whenever $\theta_+ > \|F\|_W$ since we can obtain the similar lower bound,

$$F_+ = F + \theta_+ (W' - b\mathbb{1}_C) \geq -\|F\|_W W + \theta_+ (W' - b\mathbb{1}_C) \geq (\theta_+ - \|F\|_W)W - \theta_+ b\mathbb{1}_C.$$

Moreover, in either case (10) holds for some $r > 0$, $r_0 < \infty$, since $W \in L_\infty^V$.

Hence the conclusions of Theorem 3.1 follow from Theorem 2.1. \square

3.2 Application to simulation of queues

We now return to the reflected random walk (7) to illustrate the conclusions of Theorem 2.1.

We first provide the proof of the LDP for the M/M/1 queue:

Proof of Proposition 1.1. The function $F(x) = e^{\beta x}$ is near-monotone, with $\phi = \pi(F) < \infty$. Moreover, for $\beta \in (0, |\log(\rho)|)$ the functions $V = kF$, $W = F$ solve (V3) with $k = (1 - (\alpha e^\beta + (1 - \alpha)e^{-\beta})^{-1})$. Hence the one-sided LDP follows from Theorem 2.1. The proof of the LDP for positive a , with rate function satisfying $I(a) = 0$ for $a \geq \phi$, follows as in [30, Th. 5.3] (see also the discussion of the M/M/1 queue in [26, Section 7.1].) \square

We now consider a situation in which the CLT does not hold. Consider the reflected random walk (7) with

$$\delta := -\mathbb{E}[D(k)] > 0, \quad \sigma_D^2 = \text{Var}(D) < \infty.$$

We do not assume that $\mathbb{E}[D(k)^3] < \infty$, so the CLT may not hold (see [1, Th. 3.2, Ch. 5] and [31, Chapter 17]). Moreover, Φ may not be geometrically ergodic since we do not assume that the distribution of $D(k)$ has exponential tails.

Consider first the standard estimator with $F(x) = x$ for $x \in \mathbb{R}_+$. On setting $V(x) = 1 + \frac{1}{2}\delta^{-1}(x^2 + \delta x)$ we have,

$$PV(x) = V(x) - x + R(x) \tag{32}$$

where

$$R(x) = \frac{1}{2}\delta^{-1}\sigma_D^2 - \frac{1}{2}\delta^{-1}\mathbb{E}[\{(x + D(k))^2 + \delta(x + D(k))\}\mathbb{I}(x < -D(k))].$$

Under the second-moment assumption on $D(k)$ the function R is bounded, since by Chebyshev's inequality,

$$\mathbb{E}[x^2\mathbb{I}(x < -D(k))] \leq x^2\mathbb{P}\{|D(k)| \geq x\} \leq \mathbb{E}[D(k)^2], \quad x \in \mathbb{R}_+.$$

We conclude that the assumptions of Theorem 3.1 hold, so that the standard estimator satisfies a one-sided LDP.

We now show how V can be used to construct a control-variate to obtain both upper and lower error bounds. Note that we do not know if $\pi(V) < \infty$ since we have not assumed that D possesses a third moment. Consequently, we must use some other means to establish that $\pi(H) = 0$.

Under (V3) it follows from Proposition 2.2 (i) that there exists a solution $\widehat{F} \in L_\infty^V$ to Poisson's equation (6). Moreover, it is known that the following scaling property holds,

$$\lim_{r \rightarrow \infty} r^{-2}\widehat{F}(rx) = J(x), \quad x \in \mathbb{R}_+, \tag{33}$$

where J is the fluid value function, $J(x) = \frac{1}{2}\delta^{-1}x^2$. The function \widehat{F} is convex, unique up to an additive constant, and can be chosen so that $\widehat{F}: \mathbb{X} \rightarrow \mathbb{R}_+$. The limit result (33) follows from Proposition 5.3 of [5] (see also [28, Th. 16]). Convexity is established in [4, 30] for network models.

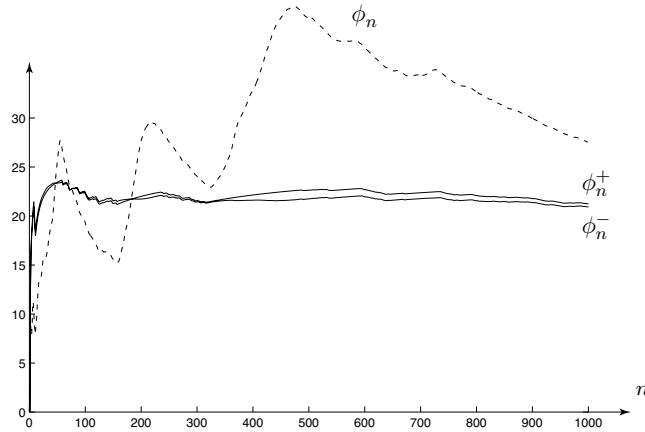


Figure 2: Monte-Carlo estimates of $\phi := \pi(F)$ with $F(x) = x$ for $x \in \mathbb{R}_+$. The stochastic process Φ is reflected random walk (7) with $\delta = -\mathbf{E}[D(k)] = 1$, and $\sigma_D^2 = 25$. The uncontrolled estimator exhibits large fluctuations around its steady-state mean. The upper and lower controlled estimators show less variability, and the bound $\phi_n^- < \phi_n^+$ is maintained throughout the run.

Iterating Poisson's equation gives,

$$P^n \widehat{F} = \widehat{F} + n\phi - \sum_{k=0}^{n-1} P^k F, \quad n \geq 1.$$

It follows from the f -Norm Ergodic Theorem [31] that for each initial condition $x \in \mathbb{R}_+$,

$$n^{-1} P^n \widehat{F}(x) = n^{-1} \widehat{F}(x) + \phi - n^{-1} \sum_{k=0}^{n-1} P^k F(x) \rightarrow 0, \quad n \rightarrow \infty.$$

The quadratic growth and positivity of \widehat{F} implies that we can find $\varepsilon > 0$ such that $\widehat{F}(x) \geq \varepsilon V(x) - 1$ for all x . Since $n^{-1} P^n \widehat{F}(x) \rightarrow 0$ as $n \rightarrow \infty$, we conclude that also $n^{-1} P^n V(x) \rightarrow 0$ as $n \rightarrow \infty$ for each x .

On setting $H = V - PV$ we see from (32) that the function H can be written

$$H(x) = x - R(x), \quad x \in \mathbb{R}_+, \quad (34)$$

where $R: \mathbf{X} \rightarrow \mathbb{R}$ is bounded. In particular, it has linear growth so that $\pi(|H|) < \infty$. Moreover, we have

$$P^n V = V - \sum_{k=0}^{n-1} P^k H, \quad n \geq 1,$$

and since $n^{-1} P^n V \rightarrow 0$ pointwise, we conclude that $\pi(H) = 0$. This justifies consideration of $F_\theta = F - \theta H$ in an asymptotically unbiased estimator of ϕ .

It also follows from the representation (34) that for any given $r \geq 1$, the set $C_{F_\theta}(r)$, is compact whenever $\theta < 1$, and $C_{-F_\theta}(r)$ is compact whenever $\theta > 1$. This structure allows the application of Theorem 2.1 to obtain confidence bounds:

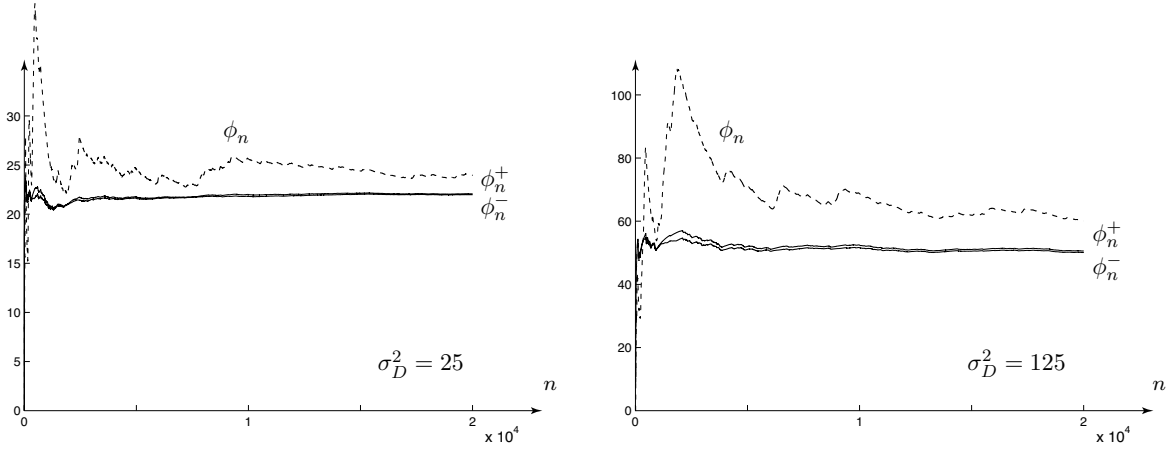


Figure 3: The plot at left shows the same simulation as shown in Figure 2, with the time horizon increased to $T = 20,000$. The plot at right shows the two controlled estimators along with the uncontrolled estimator when the variance is increased to $\sigma_D^2 = 125$. In each case the estimates obtained from the standard Monte-Carlo estimator are significantly larger than those obtained using the controlled estimator, and the bound $\phi_n^- < \phi_n^+$ again holds for all large n .

Proposition 3.2. Consider the reflected random walk (γ) with $\mathbf{E}[D(k)^2] < \infty$; $\mathbf{P}\{D(k) > 0\} > 0$; and $\delta > 0$. Let $F(x) = x$ for $x \in \mathbb{Z}_+$, fix two parameters $\theta_+ < 1$ and $\theta_- > 1$, and define the pair of estimators $\{\phi_n^-, \phi_n^+\}$ via (31) with

$$F_- = F - \theta_- H, \quad F_+ = F - \theta_+ H.$$

Then, there exists a pair of convex functions $\{I_-, I_+\}$ on \mathbb{R} , and a constant $\varepsilon_0 > 0$ such that for each initial condition $x \in \mathbf{X}$,

$$\lim_{n \rightarrow \infty} n^{-1} \log \mathbf{P}_x \{\phi_n^+ \leq c\} = -I_+(c) < 0, \quad c \in (\phi - \varepsilon_0, \phi)$$

$$\lim_{n \rightarrow \infty} n^{-1} \log \mathbf{P}_x \{\phi_n^- \geq c\} = -I_-(c) < 0, \quad c \in (\phi, \phi + \varepsilon_0).$$

Consequently, with $\{\Delta_n\}$ defined in (30), the following limit holds for each $\varepsilon \in (0, \varepsilon_0)$,

$$\begin{aligned} \lim_{n \rightarrow \infty} n^{-1} \log \mathbf{P}_x \left\{ [\phi_n - \phi] \in [-\varepsilon + \theta_+ \Delta_n, \varepsilon + \theta_- \Delta_n]^c \right\} \\ = -\min(I_+(\phi - \varepsilon), I_-(\phi + \varepsilon)) < 0. \end{aligned}$$

Proof. The assumption that $\mathbf{P}\{D(k) > 0\} > 0$ is used to deduce that π has support outside of the origin. For $\varepsilon > 0$ sufficiently small the set $C = [0, \varepsilon]$ is small, and satisfies the one-step minorization condition: for some $\delta > 0$, $P(x, \cdot) \geq \delta \nu(\cdot)$ with ν the point-mass at the origin. It follows that the functions $\{F, F_-, F_+\}$ are each non-degenerate.

The conclusions then follow from Theorem 3.1. \square

An illustration of these controlled estimators is provided in Figure 2. The sequence \mathbf{D} was chosen of the form $D(k) = A(k) - S(k)$, where \mathbf{A} and \mathbf{S} are mutually independent, i.i.d.

sequences. Given non-negative parameters μ, α, κ we set

$$\begin{aligned} \mathbb{P}\{S(k) = (1 + \kappa)\mu\} &= 1 - \mathbb{P}\{S(k) = 0\} = (1 + \kappa)^{-1} \\ \mathbb{P}\{A(k) = (1 + \kappa)\alpha\} &= 1 - \mathbb{P}\{A(k) = 0\} = (1 + \kappa)^{-1} \end{aligned}$$

Consequently, we have $\mathbb{E}[D(k)] = \mathbb{E}[A(k)] - \mathbb{E}[S(k)] = -(\mu - \alpha)$, and $\sigma_D^2 = \sigma_A^2 + \sigma_S^2 = (\mu^2 + \alpha^2)\kappa$. The simulation results shown in Figure 2 used $\mu = 4$, $\alpha = 3$, and $\kappa = 2$, so that $\delta = 1$ and $\sigma_D^2 = 25$.

The control-variate parameter values $\theta_- = 1.05$ and $\theta_+ = 1$ were used in the construction of $\{\phi_n^-, \phi_n^+\}$. While this value of θ_+ violates the strict inequality $\theta_+ > 1$ required in Proposition 3.2, we have in this case,

$$F_+(x) = x - \theta_+(x - R(x)) = R(x), \quad x \in \mathbb{R}_+.$$

The function R has mean zero and satisfies (10) when \mathbf{D} has bounded support (or just a $(2 + \epsilon)$ -moment), so Theorem 2.1 implies that the lower LDP does hold using $\{\phi_n^+\}$ when $\theta_+ = 1$.

The plot at left in Figure 3 illustrates the simulation shown previously in Figure 2, with the time horizon increased to $T = 20,000$. The plot at right shows the controlled and uncontrolled estimators with $\kappa = 5$, and hence $\sigma_D^2 = 125$. The bounds $\phi_n^- < \phi_n^+ < \phi_n$ hold for all large n even though all three estimators are asymptotically unbiased.

4 Conclusions

We have seen that it is possible to establish strong LDP asymptotics for unbounded functions even when the assumptions of Donsker and Varadhan [39, 38], or the weaker geometric ergodicity assumption are violated. We are currently developing worst-case bounds when the statistics of the process are only partially known [37], and we are also searching for ways of identifying explicit bounds on the rate function.

We are eager to develop these simulation techniques to better understand the value of the application of multiple control variates for improved confidence bounds. Applications to network models are also considered in current research.

References

- [1] S. Asmussen. *Applied Probability and Queues*. John Wiley & Sons, New York, 1987.
- [2] S. Balaji and S. P. Meyn. Multiplicative ergodicity and large deviations for an irreducible Markov chain. *Stochastic Process. Appl.*, 90(1):123–144, 2000.
- [3] E. Bolthausen. Markov process large deviations in τ -topology. *Stochastic Process. Appl.*, 25(1):95–108, 1987.
- [4] M. Chen, C. Pandit, and S. P. Meyn. In search of sensitivity in network optimization. *Queueing Systems: Theory and Applications*, 44(4):313–363, 2003.
- [5] J. G. Dai and S. P. Meyn. Stability and convergence of moments for multiclass queueing networks via fluid limit models. *IEEE Trans. Automat. Control*, 40:1889–1904, November 1995.
- [6] A. de Acosta. Large deviations for empirical measures of Markov chains. *J. Theoret. Probab.*, 3(3):395–431, 1990.
- [7] A. de Acosta and P. Ney. Large deviation lower bounds for arbitrary additive functionals of a Markov chain. *Ann. Probab.*, 26(4):1660–1682, 1998.
- [8] A. Dembo and O. Zeitouni. *Large Deviations Techniques And Applications*. Springer-Verlag, New York, second edition, 1998.
- [9] J.D. Deuschel and D.W. Stroock. *Large Deviations*. Academic Press, Boston, 1989.
- [10] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. IV. *Comm. Pure Appl. Math.*, 36(2):183–212, 1983.
- [11] P. Dupuis, R. Ellis, and A. Weiss. Large deviations for markov processes with discontinuous statistics i: general upper bounds. *Ann. Probab.*, 19(3):1280–1297, 1991.
- [12] P. Glasserman. *Monte Carlo methods in financial engineering*. Applications of Mathematics, 53. Springer, York, NY, 2004.
- [13] P. Glynn and R. Szechtman. Some new perspectives on the method of control variates. In K.T. Fang, F.J. Hickernell, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2000: Proceedings of a Conference held at Hong Kong Baptist University, Hong Kong SAR, China*, pages 27–49, Berlin, 2002. Springer-Verlag.
- [14] P. W. Glynn. Some topics in regenerative steady-state simulation. *Acta Appl. Math.*, 34(1-2):225–236, 1994.
- [15] P. W. Glynn and S. P. Meyn. A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.*, 24(2):916–931, 1996.
- [16] P. W. Glynn and D. Ormoneit. Hoeffding’s inequality for uniformly ergodic Markov chains. *Statistics and Probability Letters*, 56:143–146, 2002.

- [17] P. W. Glynn and W. Whitt. Necessary conditions for limit theorems in cumulative processes. *Stoch. Proc. Applns.*, 98:199–209, 2002.
- [18] S. G. Henderson and P. W. Glynn. Approximating martingales for variance reduction in Markov process simulation. *Math. Operations Res.*, 27(2):253–271, 2002.
- [19] S. G. Henderson, S. P. Meyn, and V. B. Tadić. Performance evaluation and policy selection in multiclass networks. *Discrete Event Dynamic Systems: Theory and Applications*, 13(1-2):149–189, 2003. Special issue on learning, optimization and decision making (invited).
- [20] S.G. Henderson. *Variance Reduction Via an Approximating Markov Process*. PhD thesis, Stanford University, Stanford, California, USA, 1997.
- [21] Arie Hordijk and Flora Spieksma. On ergodicity and recurrence properties of a Markov chain with an application to an open Jackson network. *Adv. in Appl. Probab.*, 24(2):343–376, 1992.
- [22] N.C. Jain. Large deviation lower bounds for additive functionals of Markov processes. *Ann. Probab.*, 18(3):1071–1098, 1990.
- [23] N. V. Kartashov. Criteria for uniform ergodicity and strong stability of Markov chains with a common phase space. *Theory Probab. Appl.*, 30:71–89, 1985.
- [24] N.V. Kartashov. Inequalities in theorems of ergodicity and stability for Markov chains with a common phase space. *Theory Probab. Appl.*, 30:247–259, 1985.
- [25] I. Kontoyiannis and S. P. Meyn. Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. Submitted for publication., 2003.
- [26] I. Kontoyiannis and S. P. Meyn. Spectral theory and limit theorems for geometrically ergodic Markov processes. *Ann. Appl. Probab.*, 13:304–362, 2003. Presented at the INFORMS Applied Probability Conference, NYC, July, 2001.
- [27] I. Kontoyiannis and S. P. Meyn. Finite state-space Markov chain approximations for diffusions. Submitted for publication., 2004.
- [28] P. R. Kumar and S. P. Meyn. Duality and linear programs for stability and performance analysis queueing networks and scheduling policies. *IEEE Trans. Automat. Control*, 41(1):4–17, 1996.
- [29] A. M. Law and W. D. Kelton. *Simulation Modeling and Analysis*. McGraw-Hill, New York, 3rd edition, 2000.
- [30] S. P. Meyn. Value functions, optimization, and performance evaluation in stochastic network models. Submitted for publication, 2003.
- [31] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993. online: <http://black.csl.uiuc.edu/~meyn/pages/book.html>.

- [32] S. P. Meyn and R. L. Tweedie. Computable bounds for convergence rates of Markov chains. *Ann. Appl. Probab.*, 4:981–1011, 1994.
- [33] B. L. Nelson. Control-variate remedies. *Operations Res.*, 38(4):974–992, 1990.
- [34] P. Ney and E. Nummelin. Markov additive processes. I. Eigenvalue properties and limit theorems. *Ann. Probab.*, 15(2):561–592, 1987.
- [35] P. Ney and E. Nummelin. Markov additive processes II. Large deviations. *Ann. Probab.*, 15(2):593–609, 1987.
- [36] E. Nummelin. *General Irreducible Markov Chains and Nonnegative Operators*. Cambridge University Press, Cambridge, 1984.
- [37] C. Pandit and S. P. Meyn. Extremal distributions and worst-case large-deviation bounds. Submitted for publication, 2004.
- [38] S.R.S. Varadhan. *Large Deviations and Applications*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa., 1984.
- [39] S.R.S. Varadhan. Large deviations and applications. *Expositiones Mathematicae*, 3:251–272, 1985.