http://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0030-364X, EISSN 1526-5463

Submitted to *Operations Research*

A General Framework for Importance Sampling with Markov Random Walks

Cheng-Der Fuh
Graduate Institute of Statistics, National Central University, Taoyuan City 320317, Taiwan cdffuh@gmail.com

Yanwei Jia
Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
yanweijia@cuhk.edu.hk

Steven Kou
Department of Finance, Questrom School of Business, Boston University, Boston, MA 02215, USA kou@bu.edu

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and are not intended to be a true representation of the article's final published form. Use of this template to distribute papers in print or online or to submit papers to another non-INFORM publication is prohibited.

Abstract. This is the Electronic Companion to "A General Framework for Importance Sampling with Markov Random Walks" by Cheng-Der Fuh, Yanwei Jia, and Steven Kou.

Key words: rare-event simulation, importance sampling, Markov random walks, asymptotic normal regime

Subject classifications: Simulation: systems dynamics, applications; Probability: Markov processes

Area of review: Simulation.

Electronic Companion to "A General Framework for Importance Sampling with Markov Random Walks" by Cheng-Der Fuh, Yanwei Jia, and Steven Kou

EC.1. Regularity Conditions

EC.1.1. Some Terminologies

A Markov chain $\{X_n, n \geqslant 0\}$ on a state space $\mathcal{X}$ is called *Harris recurrent* if there exist a recurrent set $\mathcal{R} \in \mathcal{B}(\mathcal{X})$, a probability measure φ on $\mathcal{R}$, a $\lambda > 0$ and an integer n_0 such that

$$P\{X_n \in \mathcal{R} \text{ for some } n \geqslant 1 | X_0 = x\} = 1, \quad P\{X_{n_0} \in A | X_0 = x\} \geqslant \lambda \varphi(A), \tag{EC.1}$$

for all $x \in \mathcal{R}$ and $A \subset \mathcal{R}$. It is known that under (EC.1), X_n admits a regenerative scheme with i.i.d. inter-regeneration times for an augmented Markov chain, which is called the "split chain;" cf. Meyn and Tweedie (2012). Assumption EC.3 guarantees that the Markov random walk (X_n, S_n) admits a similar regenerative scheme.

Denote $\varrho_{\bar{a}}(0) = \varrho_{\bar{a}}$ and let $\{\varrho_{\bar{a}}(j), j \geqslant 1\}$ be the times of consecutive visits to a recurrent state $\bar{a} \in \mathcal{X}^1$. Let $\varrho = \varrho_{\bar{a}}$ be the first time (> 0) to reach the recurrent state $\bar{a}$ of the split chain, and define $u(\alpha, \zeta) = E_\nu e^{\alpha S_\varrho - \zeta \varrho}$ for $\zeta \in \mathbb{R}$, where ν is an initial distribution on $\mathcal{X}$. Assume that

$$W := \{(\alpha, \zeta) : u(\alpha, \zeta) < \infty\} \text{ is an open subset on } \mathbb{R}^2. \tag{EC.2}$$

Ney and Nummelin (1987) shows that $D = \{\alpha : u(\alpha, \zeta) < \infty \text{ for some } \zeta\}$ is an open set and that for $\alpha \in D$, the transition kernel $\hat{P}_\alpha(x, A) = E_x\{e^{\alpha Y_1} \mathbb{1}_{\{X_1 \in A\}}\}$ has a maximal simple real eigenvalue $e^{\Lambda(\alpha)}$, where $\Lambda(\alpha)$ is the unique solution of the equation $u(\alpha, \Lambda(\alpha)) = 1$, with corresponding eigenfunction $r(x, \alpha) := E_x\left[\exp\{\alpha S_\varrho - \Lambda(\alpha)\varrho\}\right]$.

A Markov chain $\{X_n, n \geqslant 0\}$ on a state space $\mathcal{X}$ is called V-uniformly ergodic if there exist an invariant probability measure π, and a measurable function $V : \mathcal{X} \to [1, \infty)$, such that $\int V(x)\pi(\mathrm{d}x) < \infty$, and

$$\lim_{n\to\infty} \sup_{x \in \mathcal{X}} \left\{ \frac{\left|E[h(X_n)|X_0 = x] - \int h(x')\pi(\mathrm{d}x')\right|}{V(x)} : |h| \leqslant V \right\} = 0, \quad \sup_{x \in \mathcal{X}} E_x\left\{\frac{V(X_1)}{V(x)}\right\} < \infty.$$

EC.1.2. Regularity conditions

ASSUMPTION EC.1. *(i) Θ is an open set containing 0, such that for a given $\theta \in \Theta$, $\mathbb{E}[e^{\theta^\top Y_1} | X_0 = x, X_1 = x'] < \infty$ for all $x, x' \in \mathcal{X}$.*

(ii) H is an open set containing 0, such that for a given $\eta \in H$, $\mathbb{E}_x[e^{k(x, X_1, \eta)}] < \infty$ for all $x \in \mathcal{X}$.

(iii) When $\mathcal{X}$ is general (unbounded) state space, for all (θ, η) lying in a compact set, the functions $\psi(x, x', \theta), k(x, x', \eta), \phi(x, \eta)$ and their partial derivatives $\frac{\partial \psi}{\partial \theta}(x, x', \theta), \frac{\partial k}{\partial \eta}(x, x', \eta), \frac{\partial \phi}{\partial \eta}(x, \eta)$ all have linear growth in x, x'. There exist positive constant $L, C_0 > 0$ such that for any $\ell \leqslant L$, $\mathbb{E}_x\left[e^{\ell |X_1|}\right] \leqslant e^{C_0 \ell (1 + |x|)}$ for any $x \in \mathcal{X}$, and $\mathbb{E}_\nu\left[e^{\ell |X_0|}\right] < \infty$.

Assumption EC.1 (i) and (ii) are necessary to guarantee the existence of two "cumulant generating functions" and hence necessary for the feasibility of the exponential tilting. This assumption is common in exponential tilting literature. The condition (iii) is the growth and moment condition needed for (unbounded) general state Markov chain to guarantee the finiteness of quantities in our methods.

Next, we state two assumptions commonly imposed in the Markov random walk literature.

ASSUMPTION EC.2. *The underlying Markov chain* $\{X_n, n \geqslant 0\}$ *is aperiodic, irreducible and Harris recurrent (i.e.,* (EC.1) *holds).*

Assumption EC.2 is a standard assumption of the underlying Markov chain $\{X_n, n \geqslant 0\}$, cf. Ney and Nummelin (1987), which also guarantees the existence of the stationary probability measure π. The Harris recurrent condition is defined in Section EC.1.1 (cf. Meyn and Tweedie 2012).

ASSUMPTION EC.3. *When the additive part does not degenerate, i.e.,* $Y_n|X_{n-1}, X_n$ *is not deterministic, there exists a measurable function* $h : \mathcal{X} \to \mathbb{R}_+$ *and a probability measure* μ *on* $\mathcal{X} \times \mathbb{R}^d$ *such that* $\int_{A\times B} p(x, \mathrm{d}x')\rho(y|x, x')\mathrm{d}y \geqslant h(x)\mu(A \times B)$.

Assumption EC.3 is often called "minorization condition" (see, e.g., Ney and Nummelin 1987). It is a common assumption in the study of Markov random walk, and together with Assumption EC.1(i) implies that $\{(X_n, S_n), n \geqslant 0\}$ is Harris recurrent and admits a regenerative scheme. More discussions on this assumption can be found in E-Companion EC.1.1. The validity of Assumption EC.3 may affect the theoretical properties of the algorithms but not the implementation of algorithms.

The next assumption is necessary for all LD-based importance sampling methods that require the existence of the eigenvalue and eigenfunction.

ASSUMPTION EC.4. *The solution to eigenvalue-eigenfunction problem* (4) *exists. Moreover, the positive eigenfunction* $r(x, \alpha)$ *is finite and analytical.*

There are sufficient conditions to ensure Assumption EC.4; see, e.g., (EC.2) in E-Companion EC.1.1, under which, Theorem 4.1 and Lemma 4.5 in Ney and Nummelin (1987) show that the eigenvalue-eigenfunction problem (4) has a simple maximum eigenvalue whose eigenfunction is positive, finite, and analytical for x in a "full" set of $\mathcal{X}$. Sufficient conditions to extend the existence to all $x \in \mathcal{X}$ are given in Chan and Lai (2003). However, we choose to directly state it as an assumption because finding eigenvalues and eigenfunctions is not required to implement our algorithm, and we only need this assumption to establish the linkage to existing LD-based, exponential tilting methods.

ASSUMPTION EC.5. *The solution to the Poisson equation* (14) *exists.*

Assumption EC.5 is required to establish the optimal link function in Theorem 2. A sufficient condition for Assumption EC.5 is Assumption EC.2 and that the Markov chain $\{X_n, n \geqslant 0\}$ is V-uniformly ergodic

(to be defined in Section EC.1.1), cf. Theorem 17.2 in Meyn and Tweedie (2012). For the examples we considered, either one can always find an explicit solution to the Poisson equation (14), or we do not need to solve (14). Typically, Assumption EC.5 is a weaker condition than Assumption EC.4, as will be demonstrated in Lemma 3.

ASSUMPTION EC.6. *The eigenfunction in* (4) *satisfies* $\limsup_{\alpha\to 0}\mathbb{E}_\nu\left[r(X_0,\alpha)^2\right] < \infty$ *and* $\liminf_{\alpha\to 0}\inf_{x\in\mathcal{X}} r(x,\alpha) > 0$.

Assumption EC.6 is required in Theorem 3 to show the proposed importance sampling method has logarithmic efficiency under asymptotic normal regimes. Commonly used sufficient conditions for Assumption EC.6 can be found, e.g., in Collamore (2002, Remark 2.3) and Chan and Lai (2003, Example 1). Note that Assumption EC.6 is mild because only the behavior at the neighborhood of $\alpha = 0$ is needed.

EC.2. Optimal Tilting (but not practical) under Exponential Martingale for Diffusion Processes

Consider the stochastic model $\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t,\ X_0 = x_0$, and we want to estimate $\mathbb{E}_{x_0}[F(X_T)]$. Let $f(t,x) = \mathbb{E}[F(X_T)|X_t = x]$, which satisfies a partial differential equation (PDE) by the well-celebrated Feynman-Kac formula:

$$\frac{\partial f}{\partial t} + b^\top(x)\frac{\partial f}{\partial x} + \frac{1}{2}\,\mathrm{trace}(\sigma\sigma^\top\frac{\partial^2 f}{\partial x^2}) = 0,\ f(T,x) = F(x).$$

We introduce the exponential martingale defined as $\mathrm{d}L_t = L_t U_t^\top \mathrm{d}W_t,\ L_0 = 1$, where U_t determines the associated change of measure, defined as $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}|_{\mathcal{F}_T} = L_T$. U_t will be treated as a control variable. Under $\mathbb{Q}$, $W_t^{\mathbb{Q}} = W_t - \int_0^t U_s\mathrm{d}s$ is the standard Brownian motion, and $\mathrm{d}X_t = [b(X_t) + \sigma(X_t)U_t]\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t^{\mathbb{Q}}$. Then $\mathbb{E}_{x_0}[F(X_T)] = \mathbb{E}_{x_0}^{\mathbb{Q}}[F(X_T)\frac{1}{L_T}]$. Applying Itô's lemma to $f(t,X_t)\frac{1}{L_t}$, we obtain

$$F(X_T)\frac{1}{L_T} - f(X_0) = \int_0^T \frac{1}{L_t}[\frac{\partial f}{\partial x}^\top(t,X_t)\sigma(X_t) - fU_t^\top]\mathrm{d}W_t^{\mathbb{Q}}.$$

If we do importance sampling under $\mathbb{Q}$, the variance of the estimator becomes

$$\mathrm{Var}_{x_0}^{\mathbb{Q}}\left(F(X_T)\frac{1}{L_T}\right) = \mathbb{E}_{x_0}\Big[\int_0^T \frac{1}{L_t^2}|\sigma^\top(X_t)\frac{\partial f}{\partial x}(t,X_t) - fU_t|^2\mathrm{d}t\Big].$$

Hence the optimal control has the form of $U^*(t,x) = \frac{1}{f(t,x)}\sigma^\top(x)\frac{\partial f}{\partial x}(t,x)$, is a function of time t and state x only. For our approximation, we only consider it as a function of x that lies in a linear space spanned by some basis function of x.

On the other hand, for a stopping-time event associated with a first passage time $\tau_\Gamma = \inf\{t \geqslant 0 : X_t \notin \Gamma\}$ for some region Γ, the value function becomes independent of t, i.e., $f(x) = \mathbb{E}[F(X_{\tau_\Gamma})|X_t = x]$, satisfying

$$b^\top(x)\frac{\partial f}{\partial x} + \frac{1}{2}\,\mathrm{trace}(\sigma\sigma^\top\frac{\partial^2 f}{\partial x^2}) = 0, \forall x\in\Gamma,\ f(x) = F(x), \forall x\in\partial\Gamma,$$

where $\partial\Gamma$ stands for the boundary of the region Γ. The rest of the derivation is similar.

Recall in Section 4.2, the optimal g^* actually depends on the value function of the event of interest, i.e., $f(t,x)=\mathbb{E}[F(X_\tau)|X_t=x]$. In particular, $g^*(t,x)=\frac{1}{f(t,x)}\frac{\partial f}{\partial x}(t,x)$. Hence, the proposed approximation of the tilting family $B(x)\eta$ can also be viewed as an approximation to $\frac{1}{f(t,x)}\frac{\partial f}{\partial x}(t,x)$.

Given this approximation, an unbiased estimator is given by

$$\mathbb{E}_{x_0}^{\mathbb{P}_\eta}\Bigg[F(X_\tau)\exp\Bigg\{\sum_{i=1}^{\tau}-\eta^\top B^\top(X_{i-1})X_i+\eta^\top B^\top(X_{i-1})\left(X_{i-1}+b\left(X_{i-1}\right)\Delta t\right)\\+\frac{1}{2}\Delta t\eta^\top B^\top(X_{i-1})\sigma\sigma^\top(X_{i-1})B(X_{i-1})\eta\Bigg\}\Bigg].$$

The resulting tilting parameter η is chosen to minimize

$$\mathbb{E}_{x_0}^{\mathbb{P}_\eta}\Bigg[F^2(X_\tau)\exp\Bigg\{\sum_{i=1}^{\tau}-\eta^\top B^\top(X_{i-1})X_i+\eta^\top B^\top(X_{i-1})\left(X_{i-1}+b\left(X_{i-1}\right)\Delta t\right)\\+\frac{1}{2}\Delta t\eta^\top B^\top(X_{i-1})\sigma\sigma^\top(X_{i-1})B(X_{i-1})\eta\Bigg\}\Bigg].$$

EC.2.1. Conventional Small-Noise Regime and Subsolutions to a Stochastic Control Problem

The conventional regime to study the large deviation in diffusion processes is to scale "$\mathrm{d}W_t$" by "$\sqrt{\epsilon}\mathrm{d}W_t$", and to consider the SDE: $\mathrm{d}X_t=b(X_t)\mathrm{d}t+\sqrt{\epsilon}\sigma(X_t)\mathrm{d}W_t$. Suppose that we want to estimate $\epsilon\log\mathbb{E}_{x_0}[\exp\{\frac{1}{\epsilon}\log F(X_T)\}]$ and define $f(t,x)=\mathbb{E}[\exp\{\frac{1}{\epsilon}\log F(X_T)\}|X_t=x]$, which satisfies

$$\frac{\partial f}{\partial t}+b^\top(x)\frac{\partial f}{\partial x}+\frac{\epsilon}{2}\operatorname{trace}(\sigma\sigma^\top\frac{\partial^2 f}{\partial x^2})=0,\ f(T,x)=\exp\{\frac{1}{\epsilon}\log F(x)\}.$$

Then by introducing $g(t,x)=\epsilon\log f(t,x)$, we can derive that g satisfies

$$\frac{\partial g}{\partial t}+b^\top(x)\frac{\partial g}{\partial x}+\frac{1}{2}(\frac{\partial g}{\partial x})^\top\sigma\sigma^\top\frac{\partial g}{\partial x}+\frac{\epsilon}{2}\operatorname{trace}(\sigma\sigma^\top(x)\frac{\partial^2 g}{\partial x^2})=0,\ g(T,x)=F(x).$$

Its leading order is a first-order nonlinear PDE:

$$\frac{\partial u}{\partial t}+b(x)^\top\frac{\partial u}{\partial x}+\frac{1}{2}(\frac{\partial u}{\partial x})^\top\sigma\sigma^\top(x)\frac{\partial u}{\partial x}=0,\ u(T,x)=F(x). \tag{EC.3}$$

Hence, $u\approx g$. From the conclusion in the optimal change-of-measure, the optimal control in this case should be

$$U^*(t,x)=\frac{\sqrt{\epsilon}}{f(t,x)}\sigma^\top(x)\frac{\partial f}{\partial x}(t,x)=\frac{1}{\sqrt{\epsilon}}\sigma^\top(x)\frac{\partial g}{\partial x}(t,x)\approx\frac{1}{\sqrt{\epsilon}}\sigma^\top(x)\frac{\partial u}{\partial x}(t,x).$$

The existing theory on the subsolution approach to the importance sampling for diffusion processes (e.g., Fouque and Tullie 2002, Dupuis et al. 2012) assert that it suffices to find a subsolution to (EC.3) (still denoted as $u(t,x)$) and the above approximation would yield efficient importance sampling estimators.

Similarly, for a stopping-time event associated with a first passage time $\tau_\Gamma = \inf\{t \geqslant 0 : X_t \notin \Gamma\}$ for some region Γ, the value function becomes independent of t, i.e., $f(x) = \mathbb{E}[\exp\{\frac{1}{\epsilon}F(X_{\tau_\Gamma})\}|X_t = x]$, satisfying

$$b^\top(x)\frac{\partial f}{\partial x} + \frac{1}{2}\operatorname{trace}(\sigma\sigma^\top\frac{\partial^2 f}{\partial x^2}) = 0, \forall x \in \Gamma,\ f(x) = \exp\{\frac{1}{\epsilon}F(x)\}, \forall x \in \partial\Gamma.$$

Following the similar derivation, we obtain a first-order PDE:

$$b(x)^\top\frac{\partial u}{\partial x} + \frac{1}{2}(\frac{\partial u}{\partial x})^\top\sigma\sigma^\top(x)\frac{\partial u}{\partial x} = 0, \forall x \in \Gamma,\ u(x) = F(x), \forall x \in \partial\Gamma. \tag{EC.4}$$

EC.2.2. Proposed Link Function in SIRD Model

Recall that in Section 4.2, for general diffusion models (without requiring small-noise regime), we suggest the link function that equivalently approximates the tilting family by $U^*(t,x) \approx \sigma^\top(x)B(x)\eta$. This seems to imply we choose $\frac{\partial u}{\partial x}(t,x) = B(x)\eta$. Then a natural question is, is it a natural subsolution to (EC.3) or (EC.4)?

In the SIRD model in Section 6.1, for simplicity, we denote $\Sigma(x) = \sigma\sigma^\top(x)$, and it is straightforward to verify that $\Sigma(X_t)$ is strictly positive definite with probability one. We use the form $B(x) = \Sigma^{-1}(x)M(x)$, where $M(x) \in \mathbb{R}^{3\times 5}$. In particular, we notice that there exists a $\eta_0 = (\alpha, \alpha, \beta, \gamma, \beta)^\top$ such that $M(x)\eta_0 = b(x)$.

We plug-in the form $\frac{\partial u}{\partial x}(t,x) = B(x)\eta$ to PDE (EC.4), and obtain

$$b(x)^\top\frac{\partial u}{\partial x} + \frac{1}{2}(\frac{\partial u}{\partial x})^\top\sigma\sigma^\top(x)\frac{\partial u}{\partial x} = \eta_0^\top M(x)^\top\Sigma^{-1}(x)M(x)\eta + \frac{1}{2}\eta^\top M(x)^\top\Sigma^{-1}(x)M(x)\eta.$$

When $\eta = \delta\eta_0$ with $\delta \in \mathbb{R}$, we have $b(x)^\top\frac{\partial u}{\partial x} + \frac{1}{2}(\frac{\partial u}{\partial x})^\top\sigma\sigma^\top(x)\frac{\partial u}{\partial x} = \delta(1 + \frac{1}{2}\delta)\eta_0^\top M(x)^\top\Sigma^{-1}(x)M(x)\eta_0$.

Moreover, for SIRD model, the stopping time is the exit time to the region $\Gamma = \{(S,I,R) \in \mathbb{R}^3_+ : I \leqslant c\}$, and the boundary condition is $F(x) = \mathbb{1}_{\{x_2 = c\}}$, where x_2 stands for the second entry of the vector $x = (S,I,R)$. It is possible to construct a subsolution that satisfies $\frac{\partial u}{\partial x}(t,x) = B(x)\eta$ for a certain η. Since the importance sampling method only requires $\frac{\partial u}{\partial x}(t,x)$, we do not give the construction for u.

EC.3. Notions of Logarithmic Efficiency

Conventionally, in the study of rare event simulation, the properties of bounded relative error and logarithmic efficiency are often used as criteria. More specifically, let $\{A(\alpha)\}$ be a sequence of events of interest, indexed by $\alpha \in \mathbb{R}_+$ or $\alpha \in \mathbb{N}$, assume that $z(\alpha) = \mathbb{P}(A(\alpha)) \to 0$ as $\alpha \to \infty$, and for each α let $\hat{Z}(\alpha)$ be an unbiased estimator for $z(\alpha)$, i.e., $\mathbb{E}[\hat{Z}(\alpha)] = z(\alpha)$.

A sequence of estimators $\{\hat{Z}(\alpha)\}$ is called to have *logarithmic efficiency* if

$$\limsup_{\alpha\to\infty}\frac{\operatorname{Var}\left(\hat{Z}(\alpha)\right)}{z(\alpha)^{2-\varepsilon}} = 0, \tag{EC.5}$$

for all $\varepsilon > 0$. See Asmussen and Glynn (2007, Chapter 6) for more introduction to these concepts.

Note that, we often call the event "rare", because we implicitly consider the limiting behavior of a sequence of events, whose probabilities tend to zero. Only when $z(\alpha) \to 0$, the notion of the logarithmic efficiency makes sense.

EC.3.1. Motivation for Definition under the Asymptotic Normal Regime

Fixed-time events. We consider the simulation of the probability of a fixed-time event in the form of $\{S_n > c_n\}$. Under the large-deviation regime, the sequence of events of interests typically have the form of $A(n,\ell) = \{\frac{S_n}{n} - \mu > \ell\}$, indexed by n. Under the original probability $\mathbb{P}$, $\frac{S_n}{n} - \mu \to 0$ almost surely. Hence $z(n,\ell) = \mathbb{P}(A(n,\ell)) \to 0$, as $n \to \infty$ for any fixed ℓ. Therefore, it naturally fits in the framework of rare event simulation. In contrast, under the asymptotic normal regime (see Definition 2), the sequence of events of interests have the form of $A(n,\ell) = \{\sqrt{\frac{n}{\sigma^2}}(\frac{S_n}{n} - \mu) > \ell\}$, indexed by n. Under the original probability $\mathbb{P}$, $\sqrt{\frac{n}{\sigma^2}}(\frac{S_n}{n} - \mu) \to \mathcal{N}(0,1)$ in distribution for some $\sigma > 0$. Therefore, $z(n,\ell) = \mathbb{P}(A(n,\ell)) \to \Phi(-\ell) \in (0,1)$, where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution. In other words, events under the asymptotic normal regime do not directly fit into the conventional framework of rare event simulation because the limiting probability is not zero. Consequently, the concept of bounded relative error and logarithmic efficiency are not discussed in the literature of moderate-deviation simulation, e.g., Fuh and Hu (2004).

The above discussion motivates us to redefine bounded relative error and logarithmic efficiency in simulation under the asymptotic normal regime by letting $\ell \to \infty$ (after first taking limit $n \to \infty$) to have a truly rare event. More precisely, a family of estimators $\{\hat{Z}(n,\ell)\}$ is called to have *logarithmic efficiency* if

$$\limsup_{\ell\to\infty, n\to\infty, \ell/\sqrt{n}\to 0} \frac{\mathrm{Var}\left(\hat{Z}(n,\ell)\right)}{\left(\Phi(-\ell)\right)^{2-\varepsilon}} = 0,$$

for all $\varepsilon > 0$.

First-passage-time events. Finally, we examine the simulation of the probability of another typical event involving the first-passage time in the form of $\{\tau_b < T\}$. In particular, under the asymptotic normal regime, the sequence of events of interests have the form of $A(b,\ell) = \{\tau_b < \sqrt{\frac{b\sigma^2}{\mu^3}}\ell + \frac{b}{\mu}\} = \{\frac{\tau_b - \frac{b}{\mu}}{\sqrt{b\sigma^2/\mu^3}} < \ell\}$, with $\mu > 0$, $b > 0$. It is well known that $z(b,\ell) = \mathbb{P}(A(b,\ell)) \to \Phi(\ell)$ as $b \to \infty$ for any given ℓ, e.g., see Fuh (2004). In this case, we are motivated to call a family of estimators $\{\hat{Z}(b,\ell)\}$ to have *logarithmic efficiency* if

$$\limsup_{\ell\to-\infty, b\to\infty, \ell/\sqrt{b}\to 0} \frac{\mathrm{Var}\left(\hat{Z}(b,\ell)\right)}{\left(\Phi(\ell)\right)^{2-\varepsilon}} = 0,$$

for all $\varepsilon > 0$.

EC.3.2. Recovering Large Deviation Regime

Fixed-time events. We consider the following limiting regime: $n \to \infty, \ell \to \infty, \ell/\sqrt{n} \to c > 0$ for the fixed-time events. Under these regimes, we say a family of estimators $\{\hat{Z}(n,\ell)\}$ has *logarithmic efficiency* if

$$\limsup_{\ell\to\infty, n\to\infty, \ell/\sqrt{n}\to c} \frac{\mathrm{Var}\left(\hat{Z}(n,\ell)\right)}{\left(z(n,\ell)\right)^{2-\varepsilon}} = 0, \tag{EC.6}$$

for all $\varepsilon > 0$.

First-passage-time events. We consider the following limiting regime: $b \to \infty, \ell \to -\infty, \ell/\sqrt{b} \to c < 0$ for the first-passage-time events. Under these regimes, we say a family of estimators $\{\hat{Z}(b,\ell)\}$ has *logarithmic efficiency* if

$$\limsup_{\ell \to -\infty, b \to \infty, \ell/\sqrt{b} \to c} \frac{\mathrm{Var}\left(\hat{Z}(b,\ell)\right)}{\left(z(b,\ell)\right)^{2-\varepsilon}} = 0, \tag{EC.7}$$

for all $\varepsilon > 0$.

Under stronger conditions, we can show that the importance sampling estimator proposed in Theorem 2 has logarithmic efficiency in the above limiting regime. For simplicity, we make the following assumption.

ASSUMPTION EC.7. *There exists $\kappa \in \mathbb{N}$, $C_2 > C_1 > 0$, and a probability measure μ on $\mathcal{X} \times \mathbb{R}^d$ such that $C_1 \mu(A \times B) \leqslant \int_{A \times B} p(x, \mathrm{d}x') \rho(y|x,x') \mathrm{d}y \leqslant C_2 \mu(A \times B)$.*

Note that Assumption EC.7 is called the uniform recurrent conditions, which imply Assumptions EC.3 to EC.6. The reason why we need a uniform condition because the limiting point of the ratio $\ell/\sqrt{n}$ (or $\ell/\sqrt{b}$) could lie anywhere, and we still need to guarantee that the eigenfunction at this point is bounded from both below and above.

THEOREM EC.1. *Suppose Assumptions EC.1, EC.2 and EC.7 hold. Then the importance sampling estimator proposed in Theorem 2 has logarithmic efficiency in the sense of* (EC.6) *and* (EC.7).

EC.4. Calculation Details Related to the Tilting Probabilities in the Examples

EC.4.1. Heston Model

For simplicity, we denote $C = \frac{\sigma^2(1-e^{-\kappa \Delta t})}{4\kappa}$, $c = \frac{4\kappa e^{-\kappa \Delta t}}{\sigma^2(1-e^{-\kappa \Delta t})}$, and $d = \frac{4\kappa\alpha}{\sigma^2}$ in this subsection. Then

$$X_{t_{k+1}} | X_{t_k} \sim C \chi^2_d \left(c X_{t_k}\right).$$

We choose the link function as $k(x, x', \eta) = \eta x'$, and then the resulting tilting probability is

$$\frac{\mathrm{d}\mathbb{P}_{\theta,\eta}}{\mathrm{d}\mathbb{P}} = \exp\{\theta Y_{t_{k+1}} + \eta X_{t_{k+1}} - \phi(X_{t_k}, \eta) - \psi(X_{t_k}, X_{t_{k+1}}, \theta)\},$$

where

$$\phi(x,\eta) = \frac{cC\eta x}{1-2\eta C} - \frac{d}{2}\log(1-2\eta C),$$

$$\psi(x,x',\theta) = \left([\mu - \frac{1}{2}x - \frac{\rho}{\sigma}\kappa(\alpha - x)]\Delta t + \frac{\rho}{\sigma}(x' - x)\right)\theta + \frac{1}{2}x(1-\rho^2)\Delta t \theta^2$$

are two cumulant functions. More precisely, under the proposed sampling distribution $\mathbb{P}_{\theta,\eta}$, the transition probability and conditional distribution are given by

$$X_{t_{k+1}} | X_{t_k} \sim C\,\mathrm{Gamma}(\frac{d}{2}, \frac{1}{2} - \eta C, \frac{c X_{t_k}}{1-2\eta C}),$$

$$Y_{t_{k+1}}|X_{t_{k+1}},X_{t_k}\sim\mathcal{N}\left([\mu-\frac{1}{2}X_{t_k}-\frac{\rho}{\sigma}\kappa(\alpha-X_{t_k})]\Delta t+\frac{\rho}{\sigma}(X_{t_{k+1}}-X_{t_k})+\theta X_{t_k}(1-\rho^2)\Delta t, X_{t_k}(1-\rho^2)\Delta t\right),$$

where we use $\mathrm{Gamma}(\alpha,\beta,\delta)$ to denote a "non-central" Gamma distribution, which has the same distribution as $\mathrm{Gamma}(\alpha+N,\beta)$, with $N\sim\mathrm{Poisson}(\frac{\delta}{2})$.

EC.4.2. SIRD Model

Recall the notations from the main text Section 6.1. We further use b_t to denote $b(X_t)$ and similarly to $\tilde{b}_t,\sigma_t,B_t$. Moreover, let $\Sigma_t=\sigma_t\sigma_t^\top$, and $\tilde{\theta}_t=\Sigma_t^{-1}(\tilde{b}_t-b_t)=B_t\eta$. Under the alternative measure, the process is sampled from

$$\begin{pmatrix}S_{t_{k+1}}\\ I_{t_{k+1}}\\ R_{t_{k+1}}\end{pmatrix}=\begin{pmatrix}S_{t_k}-\frac{\alpha_- S_{t_k}I_{t_k}}{N_{t_k}}\Delta t\\ I_{t_k}+(\frac{\alpha_+ S_{t_k}I_{t_k}}{N_{t_k}}-\beta_- I_{t_k}-\gamma_- I_{t_k})\Delta t\\ R_{t_k}+\beta_+ I_{t_k}\Delta t\end{pmatrix}+\sqrt{\Delta t}\begin{pmatrix}-\sqrt{\frac{\alpha S_{t_k}I_{t_k}}{N_{t_k}}} & 0 & 0\\ \sqrt{\frac{\alpha S_{t_k}I_{t_k}}{N_{t_k}}} & -\sqrt{\beta I_{t_k}} & -\sqrt{\gamma I_{t_k}}\\ 0 & \sqrt{\beta I_{t_k}} & 0\end{pmatrix}\epsilon_{t_{k+1}}. \tag{EC.8}$$

The likelihood ratio is

$$\frac{\mathrm{d}\mathbb{P}^\theta_{t_k}}{\mathrm{d}\mathbb{P}_{t_k}}=\exp\{\tilde{\theta}_{t_k}^\top X_{t_{k+1}}-\tilde{\theta}_{t_k}^\top b_{t_k}-\frac{1}{2}\tilde{\theta}_{t_k}^\top\Sigma_{t_k}\tilde{\theta}_{t_k}\}=\exp\{\eta^\top B_{t_k}^\top X_{t_{k+1}}-\eta^\top B_{t_k}^\top b_{t_k}-\frac{1}{2}\eta^\top B_{t_k}^\top\Sigma_{t_k}B_{t_k}\eta\}.$$

The importance sampling estimator is

$$\mathbb{E}^{\mathbb{P}^\eta}\left[\mathbb{1}_{\{\tau_c<\infty\}}\exp\{-\sum_{k=0}^{\tau_c-1}\left(\eta^\top B_{t_k}^\top X_{t_{k+1}}-\eta^\top B_{t_k}^\top b_{t_k}-\frac{1}{2}\eta^\top B_{t_k}^\top\Sigma_{t_k}B_{t_k}\eta\right)\}\right].$$

Then the second moment of the importance sampling estimator is

$$G(\eta)=\mathbb{E}^{\mathbb{P}}\left[\mathbb{1}_{\{\tau_c<\infty\}}\exp\left\{-\sum_{k=0}^{\tau_c-1}\left(\eta^\top B_{t_k}^\top X_{t_{k+1}}-\eta^\top B_{t_k}^\top b_{t_k}-\frac{1}{2}\eta^\top B_{t_k}^\top\Sigma_{t_k}B_{t_k}\eta\right)\right\}\right].$$

The tilting parameter $\eta\in\mathbb{R}^5$ is selected by minimizing the above second moment (minimizing the variance) $G(\eta)$ of the estimator by stochastic approximation algorithm.

EC.4.3. VAR-GARCH Model

By taking $\{(y_t,H_{t+1}),t\geqslant 0\}$ as the underlying Markov chain, $\{(y_t,H_{t+1},S_t),t\geqslant 0\}$ still belongs to the general Markov random walk framework. The following is the eigenvalue problem which difficult to solve:

$$\begin{aligned}&\mathbb{E}_{y_0,H_1}[e^{\theta^\top y_1}r(y_1,H_2)]\\ =\ &\mathbb{E}[e^{\theta^\top(\mu+\rho y_0+H_1^{\frac{1}{2}}\epsilon)}r(\mu+\rho y_0+H_1^{\frac{1}{2}}\epsilon,A\odot H_1^{\frac{1}{2}}\epsilon\epsilon^\top H_1^{\frac{1}{2}}+B\odot H_1)]=e^{\Lambda(\theta)}r(y_0,H_1).\end{aligned}$$

In contrast, the Poisson equation (14) has simpler form: $\mathbb{E}_{y_0,H_1}[Y_1]-\mathbb{E}_\pi[Y_1]=\mu+\rho y-\mathbb{E}_\pi[Y_1]=\rho y_0+\mu-\mathbb{E}_\pi[Y_1]=\rho y_0+\mu-(I-\rho)^{-1}\mu=\rho y_0-\rho(I-\rho)^{-1}\mu$. The function $g(y,H)=C_1y$ with $C_1=\rho(I-\rho)^{-1}$ is a solution because

$$(I-\mathcal{P}_0)g(y,H)=C_1y-C_1(\mu+\rho y)=C_1(I-\rho)y-C_1\mu=\rho y-\rho(I-\rho)^{-1}\mu=\mathbb{E}_{y,H}[Y_1]-\mathbb{E}_\pi[Y_1]. \tag{EC.9}$$

Therefore, the solution to the Poisson equation (EC.9) for this VAR-GARCH model turns out to be a linear function of y, the first component of the Markov process (y_t, H_{t+1}). Thus, it suffices to use a linear link function in y', i.e., to take $k\left((y,H),(y',H'),\eta\right)=\eta^\top y'$.

EC.5. Verification of Regularity Conditions

We use the same notations as in E-Companion EC.4.

EC.5.1. Heston Model

We first verify Assumption EC.1. (i) holds because the additive part is conditionally normally distributed and $\psi(x,x',\theta)$ exists for all θ. (ii) holds when $\eta<\frac{1}{2C}$. (iii) holds because ψ,k,ϕ are all linear functions of (x,x'). The rest of the assumption in (iii) also follows from (ii). Assumption EC.2 holds based on the properties of the CIR process; see, e.g., Jin et al. 2013. Assumption EC.4 holds because the eigenvalue and the eigenfunction can be explicitly calculated. Assumption EC.5 holds because of Lemma 3 and Assumption EC.4.

We mimic the argument in Chan and Lai (2003, Example 2) to justify Assumption EC.3 holds. Consider two compact sets $C_X=[\epsilon,1/\epsilon]$, $C_Y=[-1/\epsilon,1/\epsilon]$, with $0<\epsilon<1$. Then for any subset $B_Y\subset C_Y$ and $x,x'\in C_X$, we have

$$\mathbb{P}(Y_1\in B_Y|X_0=x,X_1=x')=\int_{B_Y}\rho(y|x,x')\mathrm{d}y\geqslant\rho_\epsilon\,\mathrm{meas}(B_Y),$$

where $\mathrm{meas}(B_Y)$ stands for the Lebesgue measure of the set B_Y, and $\rho_\epsilon=\min_{x,x'\in C_X,y\in C_Y}\rho(y|x,x')>0$. Moreover, denote $p_\epsilon=\min_{x,x'\in C_X}p(x,x')>0$, where $p(x,x')$ is the density function of $X_1\in\mathrm{d}x'|X_0=x$.

Then

$$\begin{aligned}\mathbb{P}_x\left((X_1,Y_1)\in A\times B\right)\geqslant&\int_{A\cap C_X}p(x,x')\mathbb{P}(Y_1\in B\cap C_Y|X_0=x,X_1=x')\mathrm{d}x'\\ \geqslant&\rho_\epsilon\,\mathrm{meas}(C_Y)\frac{\mathrm{meas}(B\cap C_Y)}{\mathrm{meas}(C_Y)}\int_{A\cap C_X}p(x,x')\mathrm{d}x'\\ \geqslant&\mathbb{1}_{\{x\in C_X\}}p_\epsilon\rho_\epsilon\,\mathrm{meas}(C_X)\,\mathrm{meas}(C_Y)\frac{\mathrm{meas}(B\cap C_Y)}{\mathrm{meas}(C_Y)}\frac{\mathrm{meas}(A\cap C_X)}{\mathrm{meas}(C_X)}.\end{aligned}$$

Hence Assumption EC.3 holds with $h(x)=\mathbb{1}_{\{x\in C_X\}}p_\epsilon\rho_\epsilon\,\mathrm{meas}(C_X)\,\mathrm{meas}(C_Y)$, and $\mu(A\times B)=\frac{\mathrm{meas}(B\cap C_Y)}{\mathrm{meas}(C_Y)}\frac{\mathrm{meas}(A\cap C_X)}{\mathrm{meas}(C_X)}$.

Finally, let us verify Assumption EC.6. By the proof of Corollary 1, we know for this affine model, the eigenfunction has the form of $r(x,\theta)=e^{A(\theta)x}$, where $A(\theta)$ is the solution to an algebraic equation

$$\frac{cC(\frac{\rho}{\sigma}\theta+A)}{1-2C(\frac{\rho}{\sigma}\theta+A)}+[(-\frac{1}{2}+\frac{\rho\kappa}{\sigma})\theta\Delta t+\frac{1}{2}(1-\rho^2)\theta^2\Delta t]=\frac{\rho}{\sigma}\theta+A.$$

When $\theta=0$, the above equation reduces to $2CA^2+(cC-1)A=0$, where $cC=e^{-\kappa\Delta t}<1$ and $C>0$. (Recall that in the simulation, Δt is a fixed positive number) Hence it has a positive root. It implies that

when θ is sufficiently small, we can always find a root $A(\theta) > 0$, and hence construct an eigenfunction $r(x,\theta) = e^{A(\theta)x}$.

In this example, the initial distribution is a fixed point so $\limsup_{\theta\to 0} \mathbb{E}_\nu\left[r(X_0,\theta)^2\right] = \limsup_{\theta\to 0} e^{2A(\theta)X_0} < \infty$, i.e., the first part in Assumption EC.6 holds. Moreover, since the state space is $\mathcal{X} = \mathbb{R}_+$, $\liminf_{\alpha\to 0}\inf_{x\in\mathcal{X}} r(x,\theta) = \liminf_{\alpha\to 0}\inf_{x\geqslant 0} e^{A(\theta)x} = 1$, i.e., the second part in Assumption EC.6 holds.

EC.5.2. SIRD Model

Since the importance sampling for (21) is based on the approximation in Section 4.2, we do not need to check Assumptions.

EC.5.3. VAR-GARCH Model

We first verify Assumption EC.1. (i) holds because the additive part degenerates to $S_1 = y_1$, i.e., $\psi(y,H,y'H',\theta) = \theta^\top y'$. (ii) holds because we have chosen $k\left((y,H),(y',H'),\eta\right) = \eta^\top y'$, and $\mathbb{E}_{y_0,H_1}[e^{\eta^\top y_1}] = e^{\eta^\top(\mu+\rho y_0)+\frac{1}{2}\eta^\top H_1\eta}$. That is, $\phi(y,H,\eta) = \eta^\top(\mu+\rho y_0) + \frac{1}{2}\eta^\top H_1\eta$. (iii) holds because ψ, k, ϕ all are linear function in (y,H,y',H'). Moreover, consider $\ell_1,\ell_2 > 0$, we have

$$\mathbb{E}_{y_0,H_1}[e^{\ell_1|y_1|+\ell_2|H_2|}] \leqslant \exp\{\ell_1(|\mu|+|\rho||y_0|) + \ell_2(|W|+|B\odot H_1|)\}\mathbb{E}_{z_1\sim\mathcal{N}(0,H_1)}[e^{\ell_1|z_1|+\ell_2|z_1^\top A z_1|}] < \infty$$

as long as ℓ_1,ℓ_2 are sufficiently small. When the above is finite, its logarithm has linear growth in y_0, H_1, and ℓ_1,ℓ_2. Assumption EC.2 holds based on the properties of VAR-GARCH model (under suitable assumptions on the coefficients ρ, A, B), e.g., see Meitz and Saikkonen (2008). Assumption EC.3 holds naturally since the additive part degenerates.

Assumption EC.5 holds since we already find the solution to (14) for this model in (EC.9) explicitly.

EC.6. Two Additional Examples

We present two additional examples where the Poisson equation has an analytical solution. The first is an affine process in an autoregressive model. The second one is a finite-state-space-affine process in a Markov switching model. It is difficult to find the eigenfunctions for the second example; thus, the LD-based simulation method cannot be applied easily.

EXAMPLE EC.1. Consider a discrete-time affine model as described in Le et al. (2010). $\{X_t\in\mathbb{R}, t\geqslant 0\}$ is the underlying Markov process whose transition probability satisfies

$$\log\mathbb{E}[e^{\eta X_{t+1}}|X_t] = (\alpha+\beta X_t)\eta + \frac{1}{2}(\iota+\gamma X_t)\sigma_x^2\eta^2,$$

and the additive part $\{Y_t\in\mathbb{R}, t\geqslant 1\}$ satisfies

$$\log\mathbb{E}[e^{\theta Y_{t+1}}|X_t, X_{t+1}] = (a_1X_t + a_2X_{t+1} + a_3)\theta + \frac{1}{2}(b_1X_t + b_2X_{t+1} + b_3)^2\sigma_y^2\theta^2.$$

Then (X_t, S_t) is affine, with $\psi(x,x',\theta) = (a_1x + a_2x' + a_3)\theta + \frac{(b_1x+b_2x'+b_3)\theta^2}{2}\sigma_y^2$, $\phi(x,\eta) = (\alpha+\beta x)\eta + \frac{(\iota+\gamma x)\eta^2}{2}\sigma_x^2$. Suppose X_t admits a stationary distribution π, then under the stationary distribution, we have $\mathbb{E}_\pi[X_1] = \alpha + \beta\mathbb{E}_\pi[X_0]$. Hence $\mathbb{E}_\pi[X_0] = \frac{\alpha}{1-\beta}$. In addition, $\mathbb{E}_\pi[Y_1] = a_1\mathbb{E}_\pi[X_0] + a_2\mathbb{E}_\pi[X_1] + a_3 = \frac{\alpha(a_1+a_2)}{1-\beta} + a_3$.

Consider the Poisson equation

$$g(x) - \mathbb{E}_x[g(X_1)] = \mathbb{E}_x[Y_1] - \mathbb{E}_\pi[Y_1],$$

where $\mathbb{E}_x[Y_1] - \mathbb{E}_\pi[Y_1] = a_1x + a_2\mathbb{E}_x[X_1] - \frac{\alpha(a_1+a_2)}{1-\beta} = (a_1+a_2\beta)x - \frac{\alpha(a_1+a_2\beta)}{1-\beta}$.

We notice that the solution to the Poisson equation is also a linear function in the form of $g(x) = \bar{A}x$. Given this form, $g(x) - \mathbb{E}_x[g(X_1)] = \bar{A}x - \bar{A}(\alpha+\beta x) = \bar{A}(1-\beta)x - \bar{A}\alpha$. Therefore, $\bar{A} = \frac{a_1+a_2\beta}{1-\beta}$.

On the other hand, consider the eigenvalue problem:

$$\mathbb{E}_x[e^{\theta Y_1}r(X_1)] = e^{\Lambda(\theta)}r(x).$$

We notice that one eigenfunction is given by e^{Ax} with eigenvalue $e^{\Lambda(\theta)}$, where A depends on θ and solves a quadratic function:

$$A = a_1\theta + \frac{b_1\theta^2\sigma_y^2}{2} + \beta(A + a_2\theta + \frac{b_2\theta^2\sigma_y^2}{2}) + \frac{\gamma\sigma_x^2}{2}(A + a_2\theta + \frac{b_2\theta^2\sigma_y^2}{2})^2,$$

and $\Lambda(\theta) = a_3\theta + \frac{b_3\theta^2\sigma_y^2}{2} + \alpha(A + a_2\theta + \frac{b_2\theta^2\sigma_y^2}{2}) + \frac{\iota\sigma_x^2}{2}(A + a_2\theta + \frac{b_2\theta^2\sigma_y^2}{2})^2$.

This is because

$$\begin{aligned}
\mathbb{E}_x[e^{\theta Y_1}r(X_1)] =& \mathbb{E}_x\left[\exp\left\{AX_1 + \theta(a_1x + a_2X_1 + a_3) + \frac{\sigma_y^2\theta^2(b_1x + b_2X_1 + b_3)}{2}\right\}\right] \\
=& \mathbb{E}_x\left[\exp\left\{(A + \theta a_2 + \frac{\sigma_y^2\theta^2 b_2}{2})X_1\right\}\right] e^{(\theta a_1 + \frac{\sigma_y^2\theta^2 b_1}{2})x + \theta a_3 + \frac{\sigma_y^2\theta^2 b_3}{2}} \\
=& \exp\left\{(A + \theta a_2 + \frac{\sigma_y^2\theta^2 b_2}{2})(\alpha+\beta x) + \frac{\sigma_x^2(A + \theta a_2 + \frac{\sigma_y^2\theta^2 b_2}{2})^2}{2}(\iota+\gamma x)\right. \\
& \left. + (\theta a_1 + \frac{\sigma_y^2\theta^2 b_1}{2})x + \theta a_3 + \frac{\sigma_y^2\theta^2 b_3}{2}\right\} \\
=& \exp\left\{\left(\theta a_1 + \frac{\sigma_y^2\theta^2 b_1}{2} + \beta(A + \theta a_2 + \frac{\sigma_y^2\theta^2 b_2}{2}) + \frac{\gamma\sigma_x^2(A + \theta a_2 + \frac{\sigma_y^2\theta^2 b_2}{2})^2}{2}\right)x\right\} \\
& \times \exp\left\{\left(\theta a_3 + \frac{\sigma_y^2\theta^2 b_3}{2} + \alpha(A + \theta a_2 + \frac{\sigma_y^2\theta^2 b_2}{2}) + \frac{\iota\sigma_x^2(A + \theta a_2 + \frac{\sigma_y^2\theta^2 b_2}{2})^2}{2}\right)\right\} \\
=& e^{\Lambda(\theta) + Ax}.
\end{aligned}$$

EXAMPLE EC.2. Consider a general regime-switching $AR(1)$ model:

$$Y_{t+1} = \mu(X_{t+1}) + \beta(X_{t+1})Y_t + \sigma(X_{t+1})\epsilon_{t+1},$$

where X_t is an aperiodic, irreducible and positive recurrent Markov chain on a finite state space with transition matrix $P \in \mathbb{R}^{m\times m}$ and stationary distribution π_x, that is $\pi_x^\top P = \pi_x^\top$. We use $P[x,]$ to denote the x-th row in the transition matrix (which is a row vector). In addition, $\epsilon_{t+1} \sim \mathcal{N}(0,1)$ and is independent of other variables. Moreover, we denote μ, β, σ^2 by column vectors formed by arranging $\mu(X), \beta(X), \sigma^2(X)$ accordingly.

First, under the joint stationary distribution π, $\mathbb{E}_\pi[Y_{t+1}] = \pi_x^\top \mu + \pi_x^\top \beta \mathbb{E}_\pi[Y_{t+1}]$. Hence the stationary mean for Y is $\mathbb{E}_\pi[Y_1] = \frac{\pi_x^\top \mu}{1-\pi_x^\top \beta}$.

Consider the Poisson equation

$$g(x,y) - \mathbb{E}_{x,y}[g(X_1,Y_1)] = \mathbb{E}_{x,y}[Y_1] - \mathbb{E}_\pi[Y_1],$$

where $\mathbb{E}_{x,y}[Y_1] - \mathbb{E}_\pi[Y_1] = P[x,](\mu + \beta y) - \frac{\pi_x^\top \mu}{1-\pi_x^\top \beta}$, where $P[x,]$ denotes the x-th row of the transition matrix.

We notice that the solution to the Poisson equation has the form of $g(x,y) = \bar{g}(x) + A(x)y$, where as a vector, A satisfies $A - Pdiag\{\beta\}A = P\beta$, and $\bar{g}$ satisfies $\bar{g} - P\bar{g} = P\mu - \frac{\pi_x^\top \mu}{1-\pi_x^\top \beta} + Pdiag\{\mu\}A$. In particular, when β is a constant that does not depend on X, then we can verify that $A = (\frac{\beta}{1-\beta}, \cdots, \frac{\beta}{1-\beta})^\top$ is the solution, in which case A is also a constant that does not depend on X.

On the other hand, consider the eigenvalue problem:

$$\mathbb{E}_{x,y}[e^{\theta Y_1} r(X_1,Y_1)] = e^{\Lambda(\theta)} r(x,y).$$

When β is a constant that does not depend on X, we realize that the eigenfunction takes the form of $r(x,y) = \bar{r}(x)e^{\frac{\theta\beta}{1-\beta}y}$, and $e^{\Lambda(\theta)}$ is the largest eigenvalue associated with the matrix $Pdiag\{\exp\{\frac{\theta}{1-\beta}\mu + \frac{1}{2}\frac{\theta^2}{(1-\beta)^2}\sigma^2\}\}$, and as a vector, $\bar{r}$ is the corresponding eigenvector. However, when β is not a constant, the eigenvalue problem cannot be easily solved.

EC.7. The Example in Djehiche et al. (2022)

We review the problem studied in Djehiche et al. (2022) as follows. We modified their notations and illustrate how this problem can be embedded into our framework, and contrast our approaches through the lens of the link functions under our framework.

The original problem in Djehiche et al. (2022) concerns a (scaled by an index n) continuous-time d-dimensional pure jump process $X_t \in \mathbb{R}^d$, with $X_0 = 0$. The jump size takes values in $\{\frac{1}{n}e_1, \cdots, \frac{1}{n}e_d\}$, where $e_j = (0,\cdots,0,1,0,\cdots,0)^\top$, with only the j-th entry being 1 while others being 0; each has intensity that is proportional to $\lambda_j(X_t) := a_j(w_j - X_{t,j})e^{b\sum_{\ell=1}^d X_{t,\ell}}$, $j = 1,\cdots,d$. The total jump intensity is $R(X_t) := ne^{b\sum_{j=1}^d X_{t,j}} \sum_{j=1}^d a_j(w_j - X_{t,j})$. The event of interest is $\mathbb{P}_0(\tau_z \leqslant T)$, where $\tau_z = \inf\{t \geqslant 0 : \sum_{j=1}^d X_{t,j} \geqslant z\}$, with $\sum_{j=1}^d w_j = 1$ and $z \in (0,1)$.

Djehiche et al. (2022) find the asymptotically optimal importance sampling estimator based on the subsolution and LD principle when the model is homogeneous with $a_1 = \cdots = a_d = a^*$. In particular, Djehiche et al. (2022) suggest sampling under the distribution with $\bar{\lambda}_j(x) = \lambda_j(x)\left(1 + \frac{c}{a^*(1-\sum_{\ell=1}^d x_\ell)e^{b\sum_{\ell=1}^d x_\ell}}\right)$, and $\bar{R}(x) := R(x)\left(1 + \frac{c}{a^*(1-\sum_{\ell=1}^d x_\ell)e^{b\sum_{\ell=1}^d x_\ell}}\right)$, for some $c \in \mathbb{R}$ that is determined by numerically solving a certain equation. When the model is inhomogeneous, Djehiche et al. (2022) still suggest the same form of the sampling distribution by defining $a^* = \max_{1\leqslant \ell \leqslant d}\{a_\ell\}$, but there is no theoretical guarantee for such choices under the inhomogeneous model.

This continuous-time problem can be equivalently embedded into a discrete-time Markov random walk under our framework as follows. With abuse of notation, we denote $X_i \in \mathbb{R}^d$ with $X_0 = 0$ as a discrete-time Markov chain, where $\mathbb{P}(X_{i+1} = x + \frac{1}{n}e_j | X_i = x) = \frac{\lambda_j(x)}{\sum_{\ell=1}^d \lambda_\ell(x)}$. Moreover, we define a sequence of random variables $Y_1, \cdots, Y_i, \cdots$, such that $Y_{i+1}|X_{i+1} = x', X_i = x \sim \text{Exp}(R(x))$, i.e., an exponential distribution with intensity $R(x)$. We define $S_m = \sum_{i=1}^m Y_i$, then $\{(X_i, S_i), i \geqslant 0\}$ forms a Markov random walk. Furthermore, denote $N = \lceil nz \rceil$, and we know $\sum_{j=1}^d X_{N-1,j} = \frac{N-1}{n} < z$ and $\sum_{j=1}^d X_{N,j} = \frac{N}{n} \geqslant z$. That is, X_i takes N steps to exceed the desired threshold. Therefore, the event of interest is equivalent to $\mathbb{P}_0(S_N \leqslant T)$. Since Y_i are all positive random variables, the event of interest can also be written as $\mathbb{P}_0(\iota_T > N)$, where $\iota_T = \inf\{i \geqslant 0 : S_i > T\}$.

Under our framework, we consider the duo-exponential family for the Markov random walk $\{(X_i, S_i), i \geqslant 0\}$ through a link function $k(x, x', \eta)$. That is, the sampling transition distribution of $\{(X_i, S_i), i \geqslant 0\}$ is

$$\mathbb{P}_{\theta,\eta}(x', s'|x, s) = e^{\theta(s'-s)+k(x,x',\eta)-\psi(x,\theta)-\phi(x,\eta)}\mathbb{P}(x', s'|x, s),$$

where $e^{\psi(x,\theta)} = \frac{R(x)}{R(x)-\theta}$, and $e^{\phi(x,\eta)} = \sum_{j=1}^d \frac{\lambda_j(x)e^{k(x,x+\frac{1}{n}e_j,\eta)}}{\sum_{\ell=1}^d \lambda_\ell(x)}$.

This implies that under $\mathbb{P}_{\theta,\eta}$, $Y_{i+1}|X_{i+1} = x', X_i = x \sim \text{Exp}(R(x) - \theta)$, and $\mathbb{P}_{\theta,\eta}(X_{i+1} = x + \frac{1}{n}e_j | X_i = x) = e^{k(x,x+\frac{1}{n}e_j,\eta)-\phi(x,\eta)}\frac{\lambda_j(x)}{\sum_{\ell=1}^d \lambda_\ell(x)} \propto e^{k(x,x+\frac{1}{n}e_j,\eta)}\lambda_j(x)$. As a special case, when $d = 1$, the underlying Markov chain $\{X_i, i \geqslant 0\}$ becomes a deterministic process $X_n = \frac{i}{n}$, thus, the link function becomes unnecessary.

Note that the suggested change-of-measure by Djehiche et al. (2022) is to multiply $R(x)$ and every $\lambda_j(x)$ by the same function of x. Hence, the transition probability of $\{X_i, i \geqslant 0\}$ is unchanged. Moreover, we note that $\bar{R}(x) = R(x) + nc$ under a homogeneous model. This is equivalent to our duo-exponential family with $\theta = -nc$ and $k(x, x', \eta) = 0$, a trivial link function. Moreover, searching over c is equivalent to searching over θ while fixing the trivial link function.

In general, our approach allows us specifying a suitable link function, and we have several options. (1) Using a trivial link function, then our tilting family includes the one suggested by Djehiche et al. (2022) as a special case, and hence, optimizing over the unique tilting parameter θ should lead to superior finite-sample performance than the method by Djehiche et al. (2022) (if the computational cost and error of

the optimization stage are ignored). (2) Choosing the link function based on our Theorem 2 by solving a Poisson equation. Note that in this case, since X_i can only take finitely many values (at the grid points that are multipliers of $\frac{1}{n}$ in each dimension), the Poisson equation and the eigenvalue problem is associated with a finite-dimensional matrix, but the dimension of this matrix grows as fast as $O\left((nz)^d\right)$. Hence, it is computationally difficult to implement this procedure. (3) Other heuristic choices that do not require huge computational cost but may not have theoretical guarantee.

We implement the method in Djehiche et al. (2022) and ours based on a trivial link function, and the results are presented in Table EC.1. We can see that with an optimization stage to search for the optimal tilting parameters, our approach can slightly improve the variance reduction of the method in Djehiche et al. (2022), and the overall efficiency is close. It suggests that a trivial link function can already induce comparable performance to a sophisticated method while the potential computational costs can be tremendous if a link function is involved.

Table EC.1 Numerical results of different methods. Estimation of $\mathbb{P}_0(\iota_T > N)$, where $N = \lceil nz \rceil$, and $\iota_T = \inf\{i \geqslant 0 : S_i > T\}$. $\{(X_i, S_i), i \geqslant 0\}$ forms a Markov random walk with $\mathbb{P}(X_{i+1} = x + \frac{1}{n}e_j | X_i = x) = \frac{\lambda_j(x)}{\sum_{\ell=1}^d \lambda_\ell(x)}$, $S_m = \sum_{i=1}^m Y_i$, where $Y_1, \cdots, Y_i, \cdots$, such that $Y_{i+1}|X_{i+1} = x', X_i = x \sim \text{Exp}(R(x))$. Here, $\lambda_j(X_t) := a_j(w_j - X_{t,j})e^{b\sum_{\ell=1}^d X_{t,\ell}}$, $j = 1, \cdots, d$, and total jump intensity is $R(X_t) := ne^{b\sum_{j=1}^d X_{t,j}} \sum_{j=1}^d a_j(w_j - X_{t,j})$. In the simulation, $d = 2$, $n = 125$, $T = 5$, $b = 5$, and $a_j = 0.01$ and $w_j = 1/d$ for $j = 1, 2$. For each column, the sample size is $10^6, 10^6, 10^6, 10^7$, respectively.

	z	0.1	0.125	0.15	0.175
Plain MC	mean	7.46E-02	1.36E-02	1.93E-03	2.18E-04
	standard deviation	2.63E-04	1.16E-04	4.39E-05	4.67E-06
	elapsed time	13.43	14.13	14.73	124.42
Djehiche et al. (2022)	sd reduction ratio	2.35	4.07	7.97	16.04
	time ratio	1.44	1.36	1.47	1.91
	efficiency ratio	1.96	3.49	6.58	11.61
This paper	sd reduction ratio	2.38	4.14	8.01	18.06
(trivial link function)	time ratio	1.51	1.41	1.49	2.04
	efficiency ratio	1.93	3.48	6.56	12.63
	stage-1 time proportion	0.12	0.16	0.09	0.07

EC.8. Optimizing Tilting Parameters via the Cross-Entropy Method

In the main text Section 3.2, we suggest a two-stage algorithm for duo-exponential tilting importance sampling. The first stage is a "trial simulation" used to search for the optimal tilting parameters given a suitably chosen link function. We suggest using the stochastic gradient descent algorithm or other stochastic optimization algorithms to minimize the second moment of the importance sampling estimator $G(\theta, \eta)$ in (10), which effectively minimizes the variance of the importance sampling estimator. We describe the details for incorporating the cross-entropy (CE) method by Rubinstein and Kroese (2004) into our duo-exponential tilting framework in this part.

In short, the CE method aims to find the tilting distribution within the proposed family that is closest (in the sense of Kullback-Leibler distance) to the zero-variance distribution. For illustrative purpose, we consider the fixed-time event where τ is a deterministic time. Then the density function of the zero-variance distribution is $g^*(X_0,\cdots,X_\tau,Y_1,\cdots,Y_\tau) = \frac{1}{C}F(S_\tau)\nu(X_0)p(X_0,X_1)\cdots p(X_{\tau-1},X_\tau)\rho(Y_1|X_0,X_1),\cdots\rho(Y_\tau|X_{\tau-1},X_\tau)$, where $C=\mathbb{E}_\nu\left[F(S_\tau)\right]$ is an unknown constant. The CE method minimizes

$$\begin{aligned}\min_{\theta,\eta}&\int g^*(x_0,\cdots,x_\tau,y_1,\cdots,y_\tau)\log g^*(x_0,\cdots,x_\tau,y_1,\cdots,y_\tau)\mathrm{d}x_{0:\tau}\mathrm{d}y_{1:\tau}\\
&-\int\log\Bigg[\exp\left\{\sum_{i=1}^{\tau}\theta^\top y_i+k(x_{i-1},x_i,\eta)-\psi(x_{i-1},x_i,\theta)-\phi(x_{i-1},\eta)\right\}\\
&\qquad\nu(x_0)p(x_0,x_1)\cdots p(x_{\tau-1},x_\tau)\rho(y_1|x_0,x_1),\cdots\rho(y_\tau|x_{\tau-1},x_\tau)\Bigg]g^*(x_0,\cdots,x_\tau,y_1,\cdots,y_\tau)\mathrm{d}x_{0:\tau}\mathrm{d}y_{1:\tau}.\end{aligned}$$

The objective function reduces to

$$\begin{aligned}&\max_{\theta,\eta}\int\left[\sum_{i=1}^{\tau}\theta^\top y_i+k(x_{i-1},x_i,\eta)-\psi(x_{i-1},x_i,\theta)-\phi(x_{i-1},\eta)\right]F(S_\tau)\\
&\qquad\nu(x_0)p(x_0,x_1)\cdots p(x_{\tau-1},x_\tau)\rho(y_1|x_0,x_1),\cdots\rho(y_\tau|x_{\tau-1},x_\tau)\mathrm{d}x_{0:\tau}\mathrm{d}y_{1:\tau}\\
=&\max_{\theta,\eta}\mathbb{E}_\nu\left[F(S_\tau)\sum_{i=1}^{\tau}\left(\theta^\top Y_i+k(X_{i-1},X_i,\eta)-\psi(X_{i-1},X_i,\theta)-\phi(X_{i-1},\eta)\right)\right]\\
=&\max_{\theta,\eta}\mathbb{E}_\nu\left[F(S_\tau)\left(\theta^\top S_\tau+\sum_{i=1}^{\tau}k(X_{i-1},X_i,\eta)-\sum_{i=1}^{\tau}\psi(X_{i-1},X_i,\theta)-\sum_{i=1}^{\tau}\phi(X_{i-1},\eta)\right)\right]\\
=:&\max_{\theta,\eta}H(\theta,\eta).\end{aligned}$$

Therefore, to incorporate the CE method into our duo-exponential tilting framework, it suffices to replace $\min_{\theta,\eta}G(\theta,\eta)$ in the first stage in Algorithm 1 by $\max_{\theta,\eta}H(\theta,\eta)$. Moreover, when solving this optimization problem, one may use a previously obtained tilting parameter to perform importance sampling, that is,

$$\begin{aligned}H(\theta,\eta)=&\mathbb{E}_\nu\left[F(S_\tau)\left(\theta^\top S_\tau+\sum_{i=1}^{\tau}k(X_{i-1},X_i,\eta)-\sum_{i=1}^{\tau}\psi(X_{i-1},X_i,\theta)-\sum_{i=1}^{\tau}\phi(X_{i-1},\eta)\right)\right]\\
=&\mathbb{E}_\nu^{\mathbb{P}_{\tilde\theta,\tilde\eta}}\Bigg[F(S_\tau)\left(\theta^\top S_\tau+\sum_{i=1}^{\tau}k(X_{i-1},X_i,\eta)-\sum_{i=1}^{\tau}\psi(X_{i-1},X_i,\theta)-\sum_{i=1}^{\tau}\phi(X_{i-1},\eta)\right)\\
&\qquad e^{-\tilde\theta^\top S_\tau+\sum_{i=1}^{\tau}-k(X_{i-1},X_i,\tilde\eta)+\psi(X_{i-1},X_i,\tilde\theta)+\phi(X_{i-1},\tilde\eta)}\Bigg],\end{aligned}$$

where $(\tilde\theta,\tilde\eta)$ is any tilting parameter, and $\mathbb{P}_{\tilde\theta,\tilde\eta}$ is the distribution defined in (9). Therefore, one may use the stochastic gradient algorithm to iteratively find the desired tilting parameter. In particular, our implementation follows the suggested procedure in De Boer et al. (2005).

EC.9. Proofs

EC.9.1. Proof of Lemma 1

We provide a detailed argument for interchanging the order between the expectation and differentiation. To this end, we present the following variant of the dominated convergence theorem:

LEMMA EC.1. *Let $f_n, g_n, h_n \in L^1$ such that $g_n \geqslant 0, h_n \geqslant 0$ almost surely for all n. Suppose $f_n \to f_\infty$ almost surely, and $|f_n| \leqslant g_n + h_n$. In addition, we assume $g_n \to g_\infty$ in L^1 and almost surely, and $\mathbb{E}[h_n] \to 0$. Then $f_n \to f_\infty$ in L^1.*

Proof. By Fatou's lemma, $\mathbb{E}[|f_\infty|] \leqslant \liminf_{n\to\infty} \mathbb{E}[|f_n|] \leqslant \mathbb{E}[|g_\infty|] < \infty$, hence $f_\infty \in L^1$.

Note that $|f_n - f_\infty| \leqslant g_n + h_n + |f_\infty|$, hence we apply Fatou's lemma again to $g_n + h_n + |f_\infty| - |f_n - f_\infty| \geqslant 0$ and obtain

$$\mathbb{E}\left[\liminf_{n\to\infty} (g_n + h_n + |f_\infty| - |f_n - f_\infty|)\right] \leqslant \liminf_{n\to\infty} \mathbb{E}[g_n + h_n + |f_\infty| - |f_n - f_\infty|].$$

That is,

$$\mathbb{E}\left[g_\infty + |f_\infty| + \liminf_{n\to\infty} h_n\right] \leqslant \mathbb{E}[g_\infty] + \mathbb{E}[|f_\infty|] - \limsup_{n\to\infty} \mathbb{E}[|f_n - f_\infty|].$$

Therefore, $\limsup_{n\to\infty} \mathbb{E}[|f_n - f_\infty|] + \mathbb{E}[\liminf_{n\to\infty} h_n] \leqslant 0$. This proves $\mathbb{E}[|f_n - f_\infty|] \to 0$, i.e., $f_n \to f_\infty$ in L^1. □

The following moment estimate is also needed in the proof.

LEMMA EC.2. *Under Assumptions EC.1, EC.2, and EC.3, for a fixed n, the following statements hold.*

(i) When θ lies in the interior of Θ and sufficiently small, $\mathbb{E}_\nu\left[e^{\theta^\top S_n}|S_n|^q\right] < \infty$ for any $q \geqslant 0$.

(ii) We denote $\tilde{S}_n = \sum_{i=1}^n |X_i|$. For any ℓ sufficiently small and any $q \geqslant 0$, $\mathbb{E}_\nu[e^{\ell \tilde{S}_n} \tilde{S}_n^q] < \infty$.

(iii) When θ lies in the interior of Θ and θ, η sufficiently small, $\mathbb{E}_\nu\left[e^{\theta^\top S_n + \ell \tilde{S}_n}\left(|S_n|^q + \tilde{S}_n^p\right)\right] < \infty$ for any $p, q \geqslant 0$.

Proof.

(i) When the additive part does not degenerate, by Proposition 1(i) in Fuh (2004), under Assumptions EC.1 and EC.3, and when θ is sufficiently small, $\mathbb{E}_\nu\left[e^{|\theta||S_n|}\right] < \infty$. Hence $\mathbb{P}_\nu(|S_n| > t) \leqslant \mathbb{E}_\nu\left[e^{|\theta||S_n|}\right] e^{-|\theta|t}$. Then for any $q > 0$,

$$\mathbb{E}_\nu[|S_n|^q] = \int_0^\infty q t^{q-1} \mathbb{P}_\nu(|S_n| > t) \mathrm{d}t \leqslant q \mathbb{E}_\nu\left[e^{|\theta||S_n|}\right] \int_0^\infty t^{q-1} e^{-|\theta|t} \mathrm{d}t < \infty.$$

Since θ lies in the interior of Θ, there exists $v > 1$ such that $v\theta \in \Theta$ and $\mathbb{E}_\nu\left[e^{v|\theta||S_n|}\right] < \infty$. By Hölder's inequality,

$$\mathbb{E}_\nu\left[e^{\theta^\top S_n}|S_n|^q\right] \leqslant \left(\mathbb{E}_\nu\left[e^{v|\theta||S_n|}\right]\right)^{1/v} \left(\mathbb{E}_\nu\left[|S_n|^{qu}\right]\right)^{1/u} < \infty,$$

where $\frac{1}{v} + \frac{1}{u} = 1$.

When the additive part degenerates, it reduces to (ii).

(ii) We only need to prove the case when $\mathcal{X}$ is unbounded. When $\ell < \frac{L}{1+C_0+\cdots+C_0^n}$, by Assumption EC.1(iii), we have

$$\mathbb{E}_\nu\left[e^{\ell\tilde{S}_n}\right] = \mathbb{E}_\nu\left[e^{\ell\tilde{S}_{n-1}}\mathbb{E}_\nu\left[e^{\ell|X_n|}\middle|\mathcal{F}_{n-1}\right]\right] \leqslant e^{C_0\ell}\mathbb{E}_\nu\left[e^{\ell\tilde{S}_{n-2}+(\ell+C_0\ell)|X_{n-1}|}\right] \leqslant \cdots \leqslant e^{C}\mathbb{E}_\nu\left[e^{\ell(1+C_0+\cdots+C_0^n)|X_0|}\right],$$

for some finite constant C. Hence $\mathbb{E}_\nu\left[e^{\ell\tilde{S}_n}\right] < \infty$ for any ℓ sufficiently small. Moreover, for any $q > 0$, for the same reason, $\mathbb{E}[\tilde{S}_n^q] < \infty$. Finally, $\mathbb{E}_\nu[e^{\ell\tilde{S}_n}\tilde{S}_n^q] < \infty$ follows by Hölder's inequality.

(iii) The conclusion follows from (i) and (ii) by Hölder's inequality.

□

Fixed-time events We first consider the case of a finite-state Markov random walk.

$$G(\theta,\eta) = \mathbb{E}_\nu[\mathbb{1}_{\{S_n>c\}}e^{-\theta S_n+\sum_{i=1}^n -k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)}].$$

Fix a sufficiently small $\delta > 0$, such that $\{(\theta',\eta') : |\theta'-\theta| + |\eta'-\eta| < 2\delta\} \subset \Theta \times H$. Because X_i only takes finite many values, we denote

$$\begin{aligned} C_\delta = &\sup_{x,x',|\theta'-\theta|<\delta} |\frac{\partial}{\partial\theta}\psi(x,x',\theta')| \vee \sup_{x,x',|\eta'-\eta|<\delta} |\frac{\partial}{\partial\eta}k(x,x',\eta')| \vee \sup_{x,x',|\eta'-\eta|<\delta} |\frac{\partial}{\partial\eta}\phi(x,x',\eta')| \\ &\vee \sup_{x,x',|\theta'-\theta|<\delta} |\psi(x,x',\theta')| \vee \sup_{x,x',|\eta'-\eta|<\delta} |k(x,x',\eta')| \vee \sup_{x,x',|\eta'-\eta|<\delta} |\phi(x,x',\eta')|. \end{aligned}$$

We have

$$\begin{aligned} &\left|\sum_{i=1}^n -k(X_{i-1},X_i,\eta+\Delta\eta)+k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)\right. \\ &\qquad \left.-\psi(X_{i-1},X_i,\theta)-\phi(X_{i-1},\eta)\right| \\ \leqslant &nC_\delta(2|\Delta\eta|+|\Delta\theta|). \end{aligned}$$

We consider the case where $|\Delta\theta| \vee |\Delta\eta| < \frac{1}{4nC_\delta} \wedge \delta$. The difference has an estimate:

$$\begin{aligned} &\left|\mathbb{1}_{\{S_n>c\}}e^{-(\theta+\Delta\theta)S_n+\sum_{i=1}^n -k(X_{i-1},X_i,\eta+\Delta\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)}\right. \\ &\left.-\mathbb{1}_{\{S_n>c\}}e^{-\theta^\top S_n+\sum_{i=1}^n -k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)}\right| \\ \leqslant &\mathbb{1}_{\{S_n>c\}}e^{-\theta S_n+\sum_{i=1}^n -k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)} \\ &\times\left\{\left|e^{-\Delta\theta S_n}-1\right|e^{\sum_{i=1}^n -k(X_{i-1},X_i,\eta+\Delta\eta)+k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)-\psi(X_{i-1},X_i,\theta)-\phi(X_{i-1},\eta)}\right. \\ &\left.+\left|e^{\sum_{i=1}^n -k(X_{i-1},X_i,\eta+\Delta\eta)+k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)-\psi(X_{i-1},X_i,\theta)-\phi(X_{i-1},\eta)}-1\right|\right\} \\ \leqslant &\mathbb{1}_{\{S_n>c\}}e^{-\theta S_n+3nC_\delta}\left\{\mathbb{1}_{\{|\Delta\theta||S_n|\leqslant 1\}}C|\Delta\theta||S_n| + \mathbb{1}_{\{|\Delta\theta||S_n|>1\}}\left[e^{-\Delta\theta S_n}+1\right] + nCC_\delta(2|\Delta\eta|+|\Delta\theta|)\right\} \\ \leqslant &\tilde{C}e^{-\theta S_n}\left(|S_n|+1\right)\left(|\Delta\theta|+|\Delta\eta|\right) + \tilde{C}\mathbb{1}_{\{|\Delta\theta||S_n|>1\}}\left[e^{-(\theta+\Delta\theta)S_n}+e^{-\theta S_n}\right]. \end{aligned}$$

Hence

$$\begin{aligned}
&\frac{1}{|\Delta\theta|+|\Delta\eta|}\Big|\mathbb{1}_{\{S_n>c\}}e^{-(\theta+\Delta\theta)S_n+\sum_{i=1}^n -k(X_{i-1},X_i,\eta+\Delta\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)}\\
&-\mathbb{1}_{\{S_n>c\}}e^{-\theta^\top S_n+\sum_{i=1}^n -k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)}\Big|\\
\leqslant\ &\tilde{C}e^{-\theta S_n}\left(|S_n|+1\right)+\frac{\tilde{C}}{|\Delta\theta|}\mathbb{1}_{\{|\Delta\theta||S_n|>1\}}\left[e^{-(\theta+\Delta\theta)S_n}+e^{-\theta S_n}\right]. \qquad \text{(EC.10)}
\end{aligned}$$

To analyze (EC.10), we first note that under Assumptions EC.1, EC.2, and EC.3, Lemma EC.2 holds, then $\mathbb{E}[e^{-\theta S_n}\left(|S_n|+1\right)]<\infty$. Next we need to show that $\frac{1}{|\Delta\theta|}\mathbb{E}\left[\mathbb{1}_{\{|\Delta\theta||S_n|>1\}}\left[e^{-(\theta+\Delta\theta)S_n}+e^{-\theta S_n}\right]\right]\to 0$ as $\Delta\theta\to 0$.

Here we have

$$\frac{1}{|\Delta\theta|}\mathbb{E}\left[\mathbb{1}_{\{|\Delta\theta||S_n|>1\}}e^{-\theta S_n}\right]\leqslant\mathbb{E}\left[\mathbb{1}_{\{|\Delta\theta||S_n|>1\}}|S_n|e^{-\theta S_n}\right]\to 0$$

due to $\mathbb{E}[e^{-\theta S_n}|S_n|]<\infty$ and the monotone convergence theorem.

The last term requires to estimate the probability $\mathbb{P}\left(|S_n|>\frac{1}{|\Delta\theta|}\right)$. Under Assumptions EC.1, EC.2, and EC.3, from the proof of Lemma EC.2, it is known that $\mathbb{P}\left(|S_n|>\frac{1}{|\Delta\theta|}\right)\approx e^{-\frac{Kn}{|\Delta\theta|}}$ for some $K>0$. Then by the Hölder's inequality,

$$\frac{1}{|\Delta\theta|}\mathbb{E}\left[\mathbb{1}_{\{|S_n|>\frac{1}{|\Delta\theta|}\}}e^{-(\theta+\Delta\theta)S_n}\right]\leqslant\frac{1}{|\Delta\theta|}\left(\mathbb{P}\left(|S_n|>\frac{1}{|\Delta\theta|}\right)\right)^{1/q}\left(\mathbb{E}\left[e^{-p(\theta+\Delta\theta)S_n}\right]\right)^{1/p},$$

where $\frac{1}{p}+\frac{1}{q}=1$. We may pick $p>1$ but sufficiently close to 1 such that $-p(\theta+\Delta\theta)\in\Theta$ for any $|\Delta\theta|<\delta$. Hence $\frac{1}{|\Delta\theta|}\mathbb{E}\left[\mathbb{1}_{\{|\Delta\theta||S_n|>1\}}e^{-(\theta+\Delta\theta)S_n}\right]\to 0$ as $\Delta\theta\to 0$.

When the state space of the underlying Markov chain is unbounded, we have Assumption EC.1 (iii) on the growth condition.

Note that on the event $\{\tilde{S}_n\leqslant\frac{1}{24C_\delta(|\Delta\theta|+|\Delta\eta|)}\}$, we have

$$\begin{aligned}
&\Big|-k(X_{i-1},X_i,\eta+\Delta\eta)+k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)\\
&\qquad-\psi(X_{i-1},X_i,\theta)-\phi(X_{i-1},\eta)\Big|\\
\leqslant\ &3C_\delta(n+2\tilde{S}_n)(|\Delta\theta|\vee|\Delta\eta|)\leqslant 3nC_\delta(|\Delta\theta|\vee|\Delta\eta|)+6C_\delta\tilde{S}_n(|\Delta\theta|\vee|\Delta\eta|)\leqslant 1.
\end{aligned}$$

We consider the similar estimate on the difference:

$$
\begin{aligned}
&\Big|\mathbb{1}_{\{S_n>c\}}e^{-(\theta+\Delta\theta)S_n+\sum_{i=1}^n -k(X_{i-1},X_i,\eta+\Delta\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)}\\
&\quad-\mathbb{1}_{\{S_n>c\}}e^{-\theta^\top S_n+\sum_{i=1}^n -k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)}\Big|\\
\leqslant&\mathbb{1}_{\{S_n>c\}}e^{-\theta S_n+\sum_{i=1}^n -k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)}\\
&\quad\times\left\{\left|e^{-\Delta\theta S_n}-1\right|e^{\sum_{i=1}^n -k(X_{i-1},X_i,\eta+\Delta\eta)+k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)-\psi(X_{i-1},X_i,\theta)-\phi(X_{i-1},\eta)}\right.\\
&\quad\left.+\left|e^{\sum_{i=1}^n -k(X_{i-1},X_i,\eta+\Delta\eta)+k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)-\psi(X_{i-1},X_i,\theta)-\phi(X_{i-1},\eta)}-1\right|\right\}\\
\leqslant&\mathbb{1}_{\{S_n>c\}}e^{-\theta S_n+3nC_\delta+6C_\delta\tilde{S}_n}\left\{\mathbb{1}_{\{|\Delta\theta||S_n|\leqslant 1\}}C|\Delta\theta||S_n|+\mathbb{1}_{\{|\Delta\theta||S_n|>1\}}\left[e^{-\Delta\theta S_n}+1\right]\right\}\\
&\quad+\mathbb{1}_{\{S_n>c\}}e^{-\theta S_n+\sum_{i=1}^n -k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)}\mathbb{1}_{\{\tilde{S}_n\leqslant\frac{1}{24C_\delta(|\Delta\theta|+|\Delta\eta|)}\}}3CC_\delta(n+2\tilde{S}_n)(|\Delta\theta|+|\Delta\eta|)\\
&\quad+\mathbb{1}_{\{S_n>c\}}e^{-\theta S_n}\mathbb{1}_{\{\tilde{S}_n>\frac{1}{24C_\delta(|\Delta\theta|+|\Delta\eta|)}\}}\Big|e^{\sum_{i=1}^n -k(X_{i-1},X_i,\eta+\Delta\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)}\\
&\quad-e^{\sum_{i=1}^n -k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)}\Big|\\
\leqslant&\tilde{C}e^{-\theta S_n+6C_\delta\tilde{S}_n}\left(|S_n|+\tilde{S}_n+1\right)(|\Delta\theta|+|\Delta\eta|)+\tilde{C}\mathbb{1}_{\{|\Delta\theta||S_n|>1\}}\left[e^{-(\theta+\Delta\theta)S_n}+e^{-\theta S_n}\right]e^{6C_\delta\tilde{S}_n}\\
&\quad+\tilde{C}\mathbb{1}_{\{\tilde{S}_n>\frac{1}{24C_\delta(|\Delta\theta|+|\Delta\eta|)}\}}e^{-\theta S_n}e^{6C_\delta\tilde{S}_n}.
\end{aligned}
$$

It suffices to show

$$\mathbb{E}\left[e^{-\theta S_n+6C_\delta\tilde{S}_n}\left(|S_n|+\tilde{S}_n+1\right)\right]<\infty,$$

and when $|\Delta\theta|,|\Delta\eta|\to 0$,

$$\frac{1}{|\Delta\theta|}\mathbb{E}\left[\mathbb{1}_{\{|S_n|>\frac{1}{|\Delta\theta|}\}}\left[e^{-(\theta+\Delta\theta)S_n}+e^{-\theta S_n}\right]e^{6C_\delta\tilde{S}_n}\right]\to 0,\ \frac{1}{|\Delta\theta|+|\Delta\eta|}\mathbb{E}\left[\mathbb{1}_{\{\tilde{S}_n>\frac{1}{24C_\delta(|\Delta\theta|+|\Delta\eta|)}\}}e^{-\theta S_n}e^{6C_\delta\tilde{S}_n}\right]\to 0.$$

The first expectation is finite due to the Hölder's inequality and Lemma EC.2. The second limit holds by the same argument for the finite state-space case. The last limit holds because

$$\frac{1}{|\Delta\theta|+|\Delta\eta|}\mathbb{E}\left[\mathbb{1}_{\{\tilde{S}_n>\frac{1}{24C_\delta(|\Delta\theta|+|\Delta\eta|)}\}}e^{-\theta S_n}e^{6C_\delta\tilde{S}_n}\right]\leqslant 24C_\delta\mathbb{E}\left[\mathbb{1}_{\{\tilde{S}_n>\frac{1}{24C_\delta(|\Delta\theta|+|\Delta\eta|)}\}}\tilde{S}_ne^{-\theta S_n}e^{6C_\delta\tilde{S}_n}\right]\to 0$$

by the monotone convergence theorem and the fact $\mathbb{E}\left[\tilde{S}_ne^{-\theta S_n}e^{6C_\delta\tilde{S}_n}\right]<\infty$.

First-passage-time events To make the derivation transparent, we first consider the i.i.d. case. Note that for given $\mu>0$, the second moment has the form

$$G(\theta)=\mathbb{E}[\mathbb{1}_{\{\tau_b<T\}}e^{-\theta S_{\tau_b}+\tau_b\psi(\theta)}]=\mathbb{E}[\mathbb{1}_{\{\tau_b<T\}}e^{-\theta R(b)+\tau_b\psi(\theta)}]e^{-\theta b},$$

where $R(b)=S_{\tau_b}-b$ is the overshoot.

Define $\tilde{G}(\theta)=\mathbb{E}[\mathbb{1}_{\{\tau_b<T\}}e^{-\theta R(b)+\tau_b\psi(\theta)}]$. It suffices to show

$$\tilde{G}'(\theta)=\mathbb{E}\left[(-R(b)+\tau_b\psi'(\theta))\,\mathbb{1}_{\{\tau_b<T\}}e^{-\theta R(b)+\tau_b\psi(\theta)}\right],$$

for each θ such that $\theta, -\theta \in \Theta$.

For a given constant $C_1 > 0$, we consider $\delta > 0$ such that for any $\theta' \in B_\delta(\theta)$, $|\psi(\theta') - \psi(\theta)| \leqslant C_1|\theta' - \theta|$, and $|\psi(\theta')| < C_1|\psi(\theta)|$.

Assume $|\Delta\theta| < \delta$, and consider the difference

$$
\begin{aligned}
&\frac{1}{|\Delta\theta|}\left|\mathbb{1}_{\{\tau_b<T\}}e^{-(\theta+\Delta\theta)R(b)+\tau_b\psi(\theta+\Delta\theta)} - \mathbb{1}_{\{\tau_b<T\}}e^{-\theta R(b)+\tau_b\psi(\theta)}\right| \\
\leqslant&\frac{1}{|\Delta\theta|}\mathbb{1}_{\{\tau_b<T\}}e^{-\theta R(b)+\tau_b\psi(\theta)}\left|e^{-\Delta\theta R(b)}e^{\tau_b[\psi(\theta+\Delta\theta)-\psi(\theta)]} - 1\right| \\
\leqslant&\frac{1}{|\Delta\theta|}\mathbb{1}_{\{\tau_b<T\}}e^{-\theta R(b)+\tau_b\psi(\theta)}\left\{\left|\left(e^{-\Delta\theta R(b)} - 1\right)e^{\tau_b[\psi(\theta+\Delta\theta)-\psi(\theta)]}\right| + \left|e^{\tau_b[\psi(\theta+\Delta\theta)-\psi(\theta)]} - 1\right|\right\} \\
\leqslant&\frac{1}{|\Delta\theta|}\mathbb{1}_{\{\tau_b<T\}}e^{-\theta R(b)+\tau_b\psi(\theta)}\left\{\left|e^{-\Delta\theta R(b)} - 1\right|e^{C_1|\Delta\theta|} + \left|e^{\tau_b[\psi(\theta+\Delta\theta)-\psi(\theta)]} - 1\right|\right\} \\
\leqslant&\mathbb{1}_{\{\tau_b<T\}}e^{-\theta R(b)+\tau_b\psi(\theta)}\mathbb{1}_{\{|\Delta\theta R(b)|\leqslant 1,\ \tau_b C_1|\Delta\theta|\leqslant 1\}}\left\{R(b)e^{C_1\delta} + CC_1\tau_b\right\} \\
&+ \frac{1}{|\Delta\theta|}\mathbb{1}_{\{\tau_b<T\}}\mathbb{1}_{\{|\Delta\theta R(b)|>1,\text{ or }\tau_b C_1|\Delta\theta|>1\}}\left|e^{-(\theta+\Delta\theta)R(b)+\tau_b\psi(\theta+\Delta\theta)} + e^{-\theta R(b)+\tau_b\psi(\theta)}\right| \\
\leqslant&\mathbb{1}_{\{\tau_b<T\}}e^{-\theta R(b)+\tau_b\psi(\theta)}\left\{R(b)e^{C_1\delta} + CC_1\tau_b\right\} \\
&+ \mathbb{1}_{\{|\Delta\theta R(b)|>1,\text{ or }\tau_b C_1|\Delta\theta|>1\}}\mathbb{1}_{\{\tau_b<T\}}\frac{\left|e^{-(\theta+\Delta\theta)R(b)+\tau_b\psi(\theta+\Delta\theta)} + e^{-\theta R(b)+\tau_b\psi(\theta)}\right|}{|\Delta\theta|} \\
\leqslant&\tilde{C}e^{-\theta R(b)}\left[R(b)+1\right] + \tilde{C}\mathbb{1}_{\{|\Delta\theta R(b)|>1,\text{ or }\tau_b C_1|\Delta\theta|>1\}}\frac{\left|e^{-(\theta+\Delta\theta)R(b)} + e^{-\theta R(b)}\right|}{|\Delta\theta|}.
\end{aligned}
$$

It suffices to prove $\mathbb{E}\left[e^{-\theta R(b)}[R(b)+1]\right] < \infty$, and $\mathbb{E}\left[\mathbb{1}_{\{|\Delta\theta R(b)|>1,\text{ or }\tau_b C_1|\Delta\theta|>1\}}\frac{\left|e^{-(\theta+\Delta\theta)R(b)} + e^{-\theta R(b)}\right|}{|\Delta\theta|}\right] \to 0$ as $\Delta\theta \to 0$. For the latter, we also need to estimate the probability

$$
\begin{aligned}
&\mathbb{P}\left(|\Delta\theta R(b)| > 1,\text{ or }\tau_b C_1|\Delta\theta| > 1\right) \\
=&\mathbb{P}\left(R(b) > \frac{1}{|\Delta\theta|},\text{ or }\tau_b > \frac{1}{C_1|\Delta\theta|}\right) \\
\leqslant&\mathbb{P}\left(R(b) > \frac{1}{|\Delta\theta|}\right) + \mathbb{P}\left(\tau_b > \frac{1}{C_1|\Delta\theta|}\right) \\
\leqslant&|\Delta\theta|^2\left(\mathbb{E}[R(b)^2 + C_1^2\tau_b^2]\right).
\end{aligned}
$$

Next, we consider the case of a finite-state Markov random walk. For given $\mu > 0$,

$$
\begin{aligned}
G(\theta,\eta) &= \mathbb{E}_\nu[\mathbb{1}_{\{\tau_b<T\}}e^{-\theta S_{\tau_b}+\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)}] \\
&= \mathbb{E}_\nu[\mathbb{1}_{\{\tau_b<T\}}e^{-\theta R(b)+\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)}]e^{-\theta b}.
\end{aligned}
$$

Fix a sufficiently small $\delta > 0$, such that $\{(\theta',\eta') : |\theta' - \theta| + |\eta' - \eta| < 2\delta\} \subset \Theta \times H$. Because X_i only takes finite many values, we denote

$$
\begin{aligned}
C_\delta = &\sup_{x,x',|\theta'-\theta|<\delta}|\frac{\partial}{\partial\theta}\psi(x,x',\theta')| \vee \sup_{x,x',|\eta'-\eta|<\delta}|\frac{\partial}{\partial\eta}k(x,x',\eta')| \vee \sup_{x,x',|\eta'-\eta|<\delta}|\frac{\partial}{\partial\eta}\phi(x,x',\eta')| \\
&\vee \sup_{x,x',|\theta'-\theta|<\delta}|\psi(x,x',\theta')| \vee \sup_{x,x',|\eta'-\eta|<\delta}|k(x,x',\eta')| \vee \sup_{x,x',|\eta'-\eta|<\delta}|\phi(x,x',\eta')|.
\end{aligned}
$$

On the event $\{\tau_b < T\}$, we have

$$\Big|\sum_{i=1}^{\tau_b} -k(X_{i-1},X_i,\eta+\Delta\eta)+k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta) \\ -\psi(X_{i-1},X_i,\theta)-\phi(X_{i-1},\eta)\Big| \leqslant TC_\delta(2|\Delta\eta|+|\Delta\theta|).$$

We consider the case where $|\Delta\theta|\vee|\Delta\eta| < \frac{1}{4TC_\delta}\wedge\delta$. The difference has an estimate:

$$\begin{aligned}
&\Big|\mathbb{1}_{\{\tau_b<T\}}e^{-(\theta+\Delta\theta)R(b)+\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta+\Delta\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)} \\
&-\mathbb{1}_{\{\tau_b<T\}}e^{-\theta^\top R(b)+\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)}\Big| \\
\leqslant&\mathbb{1}_{\{\tau_b<c\}}e^{-\theta R(b)+\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)} \\
&\times\Bigg\{\left|e^{-\Delta\theta R(b)}-1\right|e^{\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta+\Delta\eta)+k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)-\psi(X_{i-1},X_i,\theta)-\phi(X_{i-1},\eta)} \\
&+\left|e^{\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta+\Delta\eta)+k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)-\psi(X_{i-1},X_i,\theta)-\phi(X_{i-1},\eta)}-1\right|\Bigg\} \\
\leqslant&\mathbb{1}_{\{\tau_b<T\}}e^{-\theta R(b)+3TC_\delta}\left\{\mathbb{1}_{\{|\Delta\theta|R(b)\leqslant 1\}}C|\Delta\theta|R(b)+\mathbb{1}_{\{|\Delta\theta|R(b)>1\}}\left[e^{-\Delta\theta R(b)}+1\right]+TCC_\delta(2|\Delta\eta|+|\Delta\theta|)\right\} \\
\leqslant&\tilde{C}e^{-\theta R(b)}\left(R(b)+1\right)\left(|\Delta\theta|+|\Delta\eta|\right)+\tilde{C}\mathbb{1}_{\{|\Delta\theta|R(b)>1\}}\left[e^{-(\theta+\Delta\theta)R(b)}+e^{-\theta R(b)}\right].
\end{aligned}$$

Hence

$$\begin{aligned}
&\frac{1}{|\Delta\theta|+|\Delta\eta|}\Big|\Big|\mathbb{1}_{\{\tau_b<T\}}e^{-(\theta+\Delta\theta)R(b)+\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta+\Delta\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)} \\
&-\mathbb{1}_{\{\tau_b<T\}}e^{-\theta R(b)+\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)}\Big| \\
\leqslant&\ \tilde{C}e^{-\theta R(b)}\left(R(b)+1\right)+\frac{\tilde{C}}{|\Delta\theta|}\mathbb{1}_{\{|\Delta\theta|R(b)>1\}}\left[e^{-(\theta+\Delta\theta)R(b)}+e^{-\theta R(b)}\right]. \qquad \text{(EC.11)}
\end{aligned}$$

To analyze (EC.11), we first note that under Assumptions EC.1, EC.2, and EC.3, by Lemma 2 in Fuh (2004), we have $\sup_{b\geqslant 0}E_\pi R(b) \leqslant E_\pi(S_1^+)^2/E_\pi S_1$, which implies that $\mathbb{E}[e^{-\theta R(b)}\left(R(b)+1\right)] < \infty$. Next, we will analyze the second term in (EC.11). By Markov inequality, we have $\frac{1}{|\Delta\theta|}\mathbb{P}\left(R(b) > \frac{1}{|\Delta\theta|}\right) \leqslant E_\pi R(b) < \infty$, via Lemma 2 in Fuh (2004). Applying Hölder's inequality, we have, under Assumptions EC.1, EC.2, and EC.3, for $\theta > 0$, $\Delta\theta > 0$, $\frac{1}{|\Delta\theta|}\mathbb{E}\left[\mathbb{1}_{\{|\Delta\theta|R(b)>1\}}\left[e^{-(\theta+\Delta\theta)R(b)}+e^{-\theta R(b)}\right]\right] \to 0$ as $\Delta\theta \to 0$.

When the state space of the underlying Markov chain is unbounded, we have Assumption EC.1 (iii) on the growth condition.

Note that on the event $\{\tau_b < T,\ \tilde{S}_T \leqslant \frac{1}{24C_\delta(|\Delta\theta|+|\Delta\eta|)}\}$, we have

$$\begin{aligned}
&\sum_{i=1}^{\tau_b}\Big|-k(X_{i-1},X_i,\eta+\Delta\eta)+k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta) \\
&\qquad-\psi(X_{i-1},X_i,\theta)-\phi(X_{i-1},\eta)\Big| \\
\leqslant&3C_\delta(T+2\tilde{S}_T)(|\Delta\theta|\vee|\Delta\eta|) \leqslant 3TC_\delta(|\Delta\theta|\vee|\Delta\eta|)+6C_\delta\tilde{S}_T(|\Delta\theta|\vee|\Delta\eta|) \leqslant 1.
\end{aligned}$$

The difference can be bounded by

$$
\begin{aligned}
&\Big|\mathbb{1}_{\{\tau_b<T\}}e^{-(\theta+\Delta\theta)R(b)+\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta+\Delta\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)}\\
&-\mathbb{1}_{\{\tau_b<T\}}e^{-\theta^\top R(b)+\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)}\Big|\\
\leqslant&\mathbb{1}_{\{\tau_b<c\}}e^{-\theta R(b)+\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)}\\
&\times\left\{\left|e^{-\Delta\theta R(b)}-1\right|e^{\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta+\Delta\eta)+k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)-\psi(X_{i-1},X_i,\theta)-\phi(X_{i-1},\eta)}\right.\\
&\left.+\left|e^{\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta+\Delta\eta)+k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)-\psi(X_{i-1},X_i,\theta)-\phi(X_{i-1},\eta)}-1\right|\right\}\\
\leqslant&\mathbb{1}_{\{\tau_b<T\}}e^{-\theta R(b)+3TC_\delta+6C_\delta\tilde{S}_T}\left\{\mathbb{1}_{\{|\Delta\theta|R(b)\leqslant 1\}}|\Delta\theta|R(b)+\mathbb{1}_{\{|\Delta\theta|R(b)>1\}}\left[e^{-\Delta\theta R(b)}+1\right]\right\}\\
&+\mathbb{1}_{\{\tau_b<T\}}e^{-\theta R(b)+\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)}\mathbb{1}_{\{\tilde{S}_T\leqslant\frac{1}{24C_\delta(|\Delta\theta|+|\Delta\eta|)}\}}3CC_\delta(T+2\tilde{S}_T)(|\Delta\theta|+|\Delta\eta|)\\
&+\mathbb{1}_{\{\tau_b<T\}}e^{-\theta R(b)}\mathbb{1}_{\{\tilde{S}_T>\frac{1}{24C_\delta(|\Delta\theta|+|\Delta\eta|)}\}}\Big|e^{\sum_{i=1}^{\tau_b}-k(X_{i-1},X_i,\eta+\Delta\eta)+\psi(X_{i-1},X_i,\theta+\Delta\theta)+\phi(X_{i-1},\eta+\Delta\eta)}\\
&-e^{\sum_{i=1}^{n}-k(X_{i-1},X_i,\eta)+\psi(X_{i-1},X_i,\theta)+\phi(X_{i-1},\eta)}\Big|\\
\leqslant&\tilde{C}e^{-\theta R(b)+6C_\delta\tilde{S}_T}\left(\tilde{S}_T+R(b)+1\right)(|\Delta\theta|+|\Delta\eta|)+\tilde{C}\mathbb{1}_{\{|\Delta\theta|R(b)>1\}}\left[e^{-(\theta+\Delta\theta)R(b)}+e^{-\theta R(b)}\right]e^{6C_\delta\tilde{S}_T}\\
&+\tilde{C}\mathbb{1}_{\{\tilde{S}_T>\frac{1}{24C_\delta(|\Delta\theta|+|\Delta\eta|)}\}}e^{-\theta R(b)+6C_\delta\tilde{S}_T}.
\end{aligned}
$$

It suffices to show

$$\mathbb{E}\left[e^{-\theta R(b)+6C_\delta\tilde{S}_T}\left(R(b)+\tilde{S}_T+1\right)\right]<\infty,$$

and when $|\Delta\theta|,|\Delta\eta|\to 0$,

$$
\begin{aligned}
&\frac{1}{|\Delta\theta|}\mathbb{E}\left[\mathbb{1}_{\{R(b)>\frac{1}{|\Delta\theta|}\}}\left[e^{-(\theta+\Delta\theta)R(b)}+e^{-\theta R(b)}\right]e^{6C_\delta\tilde{S}_T}\right]\to 0,\\
&\frac{1}{|\Delta\theta|+|\Delta\eta|}\mathbb{E}\left[\mathbb{1}_{\{\tilde{S}_T>\frac{1}{24C_\delta(|\Delta\theta|+|\Delta\eta|)}\}}e^{-\theta R(b)}e^{6C_\delta\tilde{S}_T}\right]\to 0.
\end{aligned}
$$

The first two conditions hold due to the same argument as above and the Hölder's inequality.

The last limit holds because

$$\frac{1}{|\Delta\theta|+|\Delta\eta|}\mathbb{E}\left[\mathbb{1}_{\{\tilde{S}_T>\frac{1}{24C_\delta(|\Delta\theta|+|\Delta\eta|)}\}}e^{-\theta R(b)}e^{6C_\delta\tilde{S}_T}\right]\leqslant 24C_\delta\mathbb{E}\left[\mathbb{1}_{\{\tilde{S}_T>\frac{1}{24C_\delta(|\Delta\theta|+|\Delta\eta|)}\}}\tilde{S}_Te^{-\theta R(b)}e^{6C_\delta\tilde{S}_T}\right]\to 0$$

by the monotone convergence theorem and the fact $\mathbb{E}\left[\tilde{S}_Te^{-\theta R(b)}e^{6C_\delta\tilde{S}_T}\right]<\infty$, by Lemma EC.2.

EC.9.2. Proof of Theorem 1

Under Assumption EC.4, $r(x;\alpha)$ is non-zero and well-defined.

Recall that $\{(X_n,S_n),n\geqslant 0\}$ is a Markov chain, for any classical exponential tilting probability $\mathbb{P}_{\theta*}$, the transition probability is given by (6). Moreover, based on the definition of the eigenvalue and eigenfunction in (4), we have

$$e^{\Lambda(\theta^*)}r(x,\theta^*)=\mathbb{E}_x[e^{\theta^{*\top}Y_1}r(X_1,\theta^*)]=\mathbb{E}_x\big[r(X_1,\theta^*)\mathbb{E}_x[e^{\theta^{*\top}Y_1}|X_1]\big]=\mathbb{E}_x[r(X_1,\theta^*)e^{\psi(x,X_1,\theta^*)}].$$

With $k(x,x',\eta)=\psi(x,x',\eta)+\log r(x',\eta)$, let us compute

$$\phi(x,\eta)=\log\mathbb{E}_x[e^{\psi(x,X_1,\eta)+\log r(X_1,\eta)}]=\Lambda(\eta)+\log r(x,\eta).$$

Therefore, the transition probability $\mathbb{P}_{\theta,\eta}$ defined in (9), with $\theta=\eta=\theta^*$, is

$$\mathbb{P}_{\theta^*,\theta^*}(\mathrm{d}x',\mathrm{d}s'|x,s)=e^{\log r(X_1,\theta^*)+\theta^\top(s'-s)-\Lambda(\theta^*)-\log r(x,\theta^*)}p(x,\mathrm{d}x')\rho(s'-s|x,x')\mathrm{d}s'.$$

This coincides with (6). □

EC.9.3. Proof of Corollary 1

Here, we consider the affine model and directly verify the proposed forms are the eigenvalues and the associated eigenfunction. Hence only Assumption EC.1 is required.

Under the affine structure, $r(x,\theta)=A(\theta)^\top x$ is an eigenfunction if and only if

$$\begin{aligned}&E_x[e^{\theta^\top Y_1}r(X_1,\theta)]=\mathbb{E}_x\big[r(X_1,\theta)\mathbb{E}_x[e^{\theta^\top Y_1}|X_1]\big]=\mathbb{E}[e^{A(\theta)^\top X_1+D_0(\theta)^\top X_1+D_1(\theta)^\top x+D_2(\theta)}]\\ =&e^{\big[C_1\big(A(\theta)+D_0(\theta)\big)+D_1(\theta)\big]^\top x+C_2\big(A(\theta)+D_0(\theta)\big)+D_2(\theta)}=e^{\Lambda(\theta)}e^{A(\theta)^\top x}.\end{aligned}$$

This is equivalent to $C_1\big(A(\theta)+D_0(\theta)\big)+D_1(\theta)=A(\theta)$ and $C_2\big(A(\theta)+D_0(\theta)\big)+D_2(\theta)=\Lambda(\theta)$. For any given θ, $D_0(\theta),D_1(\theta),D_2(\theta)$ are three constants, and $C_1(\cdot),C_2(\cdot)$ are two functions. Solving this system of algebraic equations leads to the eigenfunction and eigenvalue.

Moreover, in this case, the classical exponential tilting family is

$$\mathbb{P}_\theta(\mathrm{d}x',\mathrm{d}s'|x,s)=e^{A(\theta)^\top(x'-x)-\Lambda(\theta)+\theta^\top(s'-s)}p(x,\mathrm{d}x')\rho(s'-s|x,x')\mathrm{d}s'.$$

On the other hand, the duo-exponential tilting distribution with link function $k(x,x',\eta)=\eta^\top x'$ is

$$\mathbb{P}_{\theta,\eta}(\mathrm{d}x',\mathrm{d}s'|x,s)=e^{\eta^\top x'+\theta^\top(s'-s)-C_1(\eta)^\top x-C_2(\eta)-D_0(\theta)^\top x'-D_1(\theta)^\top x-D_2(\theta)}p(x,\mathrm{d}x')\rho(s'-s|x,x')\mathrm{d}s'.$$

By direct comparison, we could see that when $\eta=A(\theta)+D_0(\theta)$, probability $\mathbb{P}_{\theta,\eta}(\mathrm{d}x',\mathrm{d}s'|x,s)$ coincides with $\mathbb{P}_\theta(\mathrm{d}x',\mathrm{d}s'|x,s)$. Therefore, the proposed duo-exponential tilting family contains the classical exponential tilting as a proper subset. □

EC.9.4. Proof of Lemma 2

First, under Assumption EC.1, we claim $\psi(x,x',\theta)$ and $\phi(x,\eta)$ are convex functions of θ and η, respectively, and hence Θ and H are both convex sets. This is a well-known result for cumulant generating functions. For example, it can be proved by Hölder's inequality: for given $\iota>0$

$$\begin{aligned}e^{\phi(x,\iota\eta_1+(1-\iota)\eta_2)}=&\mathbb{E}_x[e^{(\iota\eta_1+(1-\iota)\eta_2)^\top\tilde{k}(x,X_1)}]=\mathbb{E}_x[(e^{\eta_1^\top\tilde{k}(x,X_1)})^\iota(e^{\eta_2^\top\tilde{k}(x,X_1)})^{1-\iota}]\\ \leqslant&\big(\mathbb{E}_x[e^{\eta_1^\top\tilde{k}(x,X_1)}]\big)^\iota\big(\mathbb{E}_x[e^{\eta_2^\top\tilde{k}(x,X_1)}]\big)^{1-\iota}=e^{\iota\phi(x,\eta_1)+(1-\iota)\phi(x,\eta_2)}.\end{aligned}$$

Hence $\phi\big(\iota\eta_1+(1-\iota)\eta_2,x\big)\leqslant\iota\phi(x,\eta_1)+(1-\iota)\phi(x,\eta_2)$. That is, ϕ is convex in η.

Therefore, inside the expectation, for any trajectory, $F(X_\tau,S_\tau)\exp\{-\theta^\top S_\tau+\sum_{i=1}^\tau-\eta^\top\tilde{k}(X_{i-1},X_i)+\phi(X_{i-1},\eta)+\psi(X_{i-1},X_i,\eta)\}$ is a composition between exponential function and a convex function and hence remains convex in terms of (θ,η). After taking the expectation, $G(\theta,\eta)$ is also convex. □

EC.9.5. Proof of Theorem 2

We first present a lemma about the optimal log-likelihood ratio for a joint normal distribution.

LEMMA EC.3. *Under the LAN family and asymptotic normal regime, the second moment of the Monte-Carlo estimator* (10) *is asymptotically equivalent to* $\mathbb{E}[\bar{F}^2(N)e^{-N_{\tilde{\mathbb{P}}}}]$*, where* $(N, N_{\tilde{\mathbb{P}}})$ *follows a joint normal distribution. Suppose*

$$\begin{pmatrix} N \\ N_{\tilde{\mathbb{P}}} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_N \\ \mu_{\tilde{\mathbb{P}}} \end{pmatrix}, \begin{pmatrix} \Sigma_N & \Sigma_{LN}^\top \\ \Sigma_{LN} & \sigma_{\tilde{\mathbb{P}}}^2 \end{pmatrix} \right).$$

Then the minimum of $\mathbb{E}[\bar{F}^2(N)e^{-N_{\tilde{\mathbb{P}}}}]$ *is achieved when* $\sigma_{\tilde{\mathbb{P}}}^2 = \Sigma_{LN}^\top \Sigma_N^{-1} \Sigma_{LN}$.

Proof. First, we note that under the LAN family assumption of the likelihood ratio and asymptotic normality of the statistic of interest $\bar{F}^2(S_\tau)$, by using a standard joint central limit theorem, we have that the second moment of the Monte-Carlo estimator (10) is asymptotically equivalent to $\mathbb{E}[\bar{F}^2(N)e^{-N_{\tilde{\mathbb{P}}}}]$. (3.6) and (3.7) in Fuh and Hu (2007) presents the simple case in finite state Markov chains. Second, since $N_{\tilde{\mathbb{P}}}$ is the limit of a log-likelihood ratio, $\mathbb{E}[e^{N_{\tilde{\mathbb{P}}}}] = 1$. It means $\mu_{\tilde{\mathbb{P}}} = -\frac{1}{2}\sigma_{\tilde{\mathbb{P}}}^2$.

We calculate the objective by conditioning on N first. Since

$$N_{\tilde{\mathbb{P}}}|N \sim \mathcal{N}\left(\mu_{\tilde{\mathbb{P}}} + \Sigma_{LN}^\top \Sigma_N^{-1}(N - \mu_N), \sigma_{\tilde{\mathbb{P}}}^2 - \Sigma_{LN}^\top \Sigma_N^{-1} \Sigma_{LN}\right),$$

we have

$$\mathbb{E}[\bar{F}^2(N)e^{-N_{\tilde{\mathbb{P}}}}] = \mathbb{E}[\bar{F}^2(N)\mathbb{E}[e^{-N_{\tilde{\mathbb{P}}}}|N]] = \mathbb{E}[\bar{F}^2(N)e^{-\Sigma_{LN}^\top \Sigma_N^{-1}(N-\mu_N)}]e^{\sigma_{\tilde{\mathbb{P}}}^2 - \frac{1}{2}\Sigma_{LN}^\top \Sigma_N^{-1}\Sigma_{LN}}.$$

Hence it is a decreasing function in $\sigma_{\tilde{\mathbb{P}}}^2$, and the joint normality implies that $\sigma_{\tilde{\mathbb{P}}}^2 \geqslant \Sigma_{LN}^\top \Sigma_N^{-1} \Sigma_{LN}$. This proves our claim. Moreover, when $\sigma_{\tilde{\mathbb{P}}}^2 = \Sigma_{LN}^\top \Sigma_N^{-1} \Sigma_{LN}$, there exist a vector $\bar{\theta}$ and a scalar c, such that $N_{\tilde{\mathbb{P}}} = \bar{\theta}^\top N + c$. □

Now let us first prove an intermediary result. We use the following notion to denote the log-likelihood ratio up to time n induced by our proposed change of measure (9):

$$LR_n(\theta, \eta) := \sum_{i=1}^{n} \left[\theta^\top Y_i + k(X_{i-1}, X_i, \eta) - \psi(X_{i-1}, X_i, \theta) - \phi(X_{i-1}, \eta)\right].$$

LEMMA EC.4. *Under Assumptions EC.1, EC.2, and EC.5, the asymptotic optimal tilting distribution within the LAN family for an event in the asymptotic normal regime is achieved by choosing the link function according to* $k(x, x', \eta) = \psi(x, x', \eta) + \eta^\top g(x')$*, where* g *satisfies* (14)*, and the tilting parameter satisfies* $\theta = O(\frac{1}{\sqrt{n}})$*,* $\eta = O(\frac{1}{\sqrt{n}})$*, and* $\theta - \eta = o(\frac{1}{\sqrt{n}})$.

Proof. Assumption EC.1 guarantees the existence of ψ and Assumption EC.5 assumes the existence of g.

Given this link function $k(x, x', \eta) = \psi(x, x', \eta) + \eta^\top g(x')$, based on the definition, the cumulant function is

$$e^{\phi(x,\eta)} = \mathbb{E}[e^{k(X_0, X_1, \eta)}|X_0 = x] = \mathbb{E}[e^{\psi(X_0, X_1, \eta) + \eta^\top g(X_1)}|X_0 = x] = \mathbb{E}[e^{\eta^\top Y_1 + \eta^\top g(X_1)}|X_0 = x].$$

Therefore, $\frac{\partial\phi}{\partial\eta}(x,0)=\mathbb{E}_x[Y_1+g(X_1)]$. Since g solves (14), $g(x)-\mathbb{E}_x[g(X_1)]=\mathbb{E}_x[Y_1]-\mathbb{E}_\pi[Y_1]$, we obtain $\frac{\partial\phi}{\partial\eta}(x,0)=g(x)+\mathbb{E}_\pi[Y_1]$.

Now let us check the log-likelihood ratio when $\theta=\eta=-\frac{\bar\theta}{\sqrt{n}}$:

$$\begin{aligned}
LR_n(-\frac{\bar\theta}{\sqrt{n}},-\frac{\bar\theta}{\sqrt{n}}) &= -\frac{\bar\theta^\top}{\sqrt{n}}\sum_{i=1}^n[Y_i+g(X_i)]-\sum_{i=1}^n\phi(X_{i-1},-\frac{\bar\theta}{\sqrt{n}})\\
&= -\frac{\bar\theta^\top}{\sqrt{n}}\sum_{i=1}^n[Y_i+g(X_i)-\frac{\partial\phi}{\partial\eta}(X_{i-1},0)]-\frac{1}{2n}\sum_{i=1}^n\bar\theta^\top\frac{\partial^2\phi}{\partial\eta^2}(X_{i-1},0)\bar\theta+o(1)\\
&= -\frac{\bar\theta^\top}{\sqrt{n}}\sum_{i=1}^n\big[Y_i+g(X_i)-g(X_{i-1})-\mathbb{E}_\pi[Y_1]\big]-\frac{1}{2n}\sum_{i=1}^n\bar\theta^\top\frac{\partial^2\phi}{\partial\eta^2}(X_{i-1},0)\bar\theta+o(1)\\
&= -\frac{\bar\theta^\top}{\sqrt{n}}\sum_{i=1}^n\big[Y_i-\mathbb{E}_\pi[Y_1]\big]-\frac{\bar\theta^\top}{\sqrt{n}}[g(X_n)-g(X_0)]-\frac{1}{2n}\sum_{i=1}^n\bar\theta^\top\frac{\partial^2\phi}{\partial\eta^2}(X_{i-1},0)\bar\theta+o(1)\\
&= \underbrace{-\frac{\bar\theta^\top}{\sqrt{n}}\big(S_n-n\mathbb{E}_\pi[Y_1]\big)}_{\text{central limit theorem}}\underbrace{-\frac{\bar\theta^\top}{\sqrt{n}}[g(X_n)-g(X_0)]}_{\to 0}\underbrace{-\frac{1}{2n}\sum_{i=1}^n\bar\theta^\top\frac{\partial^2\phi}{\partial\eta^2}(X_{i-1},0)\bar\theta}_{\text{law of large numbers}}+o(1).
\end{aligned}$$

By the central limit theorem in Theorem 17.5.3 of Meyn and Tweedie (2012), Assumption EC.2, and Slutsky's theorem, $LR_n(-\frac{\bar\theta}{\sqrt{n}},-\frac{\bar\theta}{\sqrt{n}})$ converges to a normal distribution.

Next, we check the asymptotic mean and variance of the log likelihood ratio. For simplicity, we denote by $m\in\mathbb{R}^d, C\in\mathbb{R}^{d\times d}$ the asymptotic mean and variance of S_n. In particular, $m=\mathbb{E}_\pi[Y_1]$, $C=\lim_{n\to 0^+}\frac{1}{n}\operatorname{Var}_\pi(S_n)$. The asymptotic mean of $LR_n(-\frac{\bar\theta}{\sqrt{n}},-\frac{\bar\theta}{\sqrt{n}})$ is determined by the law of large numbers term that converges to $-\frac{1}{2}\bar\theta^\top\mathbb{E}_\pi[\frac{\partial^2\phi}{\partial\eta^2}(X_0,0)]\bar\theta$. The asymptotic variance of $LR_n(\bar\theta,\bar\theta)$ is determined by the central limit theorem term, which equals to $\bar\theta^\top C\bar\theta$. It remains to verify $C=\mathbb{E}_\pi[\frac{\partial^2\phi}{\partial\eta^2}(X_0,0)]$.

Denote $M_n=S_n+g(X_n)-nm-g(X_0)=\sum_{i=1}^n z_i$, with $z_n=Y_n+g(X_n)-g(X_{n-1})-m$. Notice that M_n is a martingale under any $\mathbb{P}_x$, and $\mathbb{E}_x[M_n]=0$. Hence $\operatorname{Var}_x(\frac{1}{\sqrt{n}}M_n)=\frac{1}{n}\sum_{i=1}^n\mathbb{E}_x[z_iz_i^\top]=\frac{1}{n}\sum_{i=1}^n\mathbb{E}_x[\frac{\partial^2\phi}{\partial\eta^2}(X_{i-1},0)]$. Recall that $\frac{1}{\sqrt{n}}M_n\to\mathcal{N}(m,C)$ in distribution under any $\mathbb{P}_x$. Therefore $\operatorname{Var}_x(\frac{1}{\sqrt{n}}M_n)\to C$ as $n\to\infty$ for any x. In particular, we take x follows the stationary distribution and obtain $\mathbb{E}_\pi[\frac{\partial^2\phi}{\partial\eta^2}(X_0,0)]=C$. This proves that our tilting family also belongs to LAN family.

Moreover, this normal distribution is a linear transformation of S_n. Therefore, by Lemma EC.3, the proposed tilting probability is asymptotically optimal within the LAN family. This is because the joint normal distributions for the log-likelihood ratio and the event considered in Lemma EC.3 are exhaustive for the limiting distributions of all twisted probabilities in the LAN family (see Definitions 1 and 2). Lemma EC.3 shows that the proposed tilting probability minimizes the asymptotic variance and hence is asymptotically optimal within the LAN family.

Lastly, we relax the condition to $\theta = -\frac{\bar{\theta}_n}{\sqrt{n}}$, $\eta = -\frac{\bar{\eta}_n}{\sqrt{n}}$, and $\lim_{n\to\infty} \bar{\theta}_n = \lim_{n\to\infty} \bar{\eta}_n = \bar{\theta} < \infty$. In this case, the log-likelihood ratio becomes

$$
\begin{aligned}
&LR_n(-\frac{\bar{\theta}_n}{\sqrt{n}}, -\frac{\bar{\eta}_n}{\sqrt{n}}) \\
=&-\frac{\bar{\eta}_n^\top}{\sqrt{n}}\sum_{i=1}^n [Y_i + g(X_i)] - \sum_{i=1}^n \phi(X_{i-1}, -\frac{\bar{\eta}_n}{\sqrt{n}}) + \sum_{i=1}^n [-\frac{\bar{\theta}_n^\top - \bar{\eta}_n^\top}{\sqrt{n}} Y_i - \psi(X_{i-1}, X_i, -\frac{\bar{\theta}_n}{\sqrt{n}}) + \psi(X_{i-1}, X_i, -\frac{\bar{\eta}_n}{\sqrt{n}})] \\
=&-\frac{\bar{\eta}_n^\top}{\sqrt{n}}\sum_{i=1}^n [Y_i + g(X_i)] - \sum_{i=1}^n \phi(X_{i-1}, -\frac{\bar{\eta}_n}{\sqrt{n}}) - \frac{\bar{\theta}_n^\top - \bar{\eta}_n^\top}{\sqrt{n}}\sum_{i=1}^n [Y_i - \frac{\partial \psi}{\partial \theta}(X_{i-1}, X_i, 0)] \\
&-\frac{1}{2}\sum_{i=1}^n [\frac{1}{n}\bar{\theta}_n^\top \frac{\partial^2 \psi}{\partial \theta^2}(X_{i-1}, X_i, 0)\bar{\theta}_n - \frac{1}{n}\bar{\eta}_n^\top \frac{\partial^2 \psi}{\partial \theta^2}(X_{i-1}, X_i, 0)\bar{\eta}_n + o(\frac{1}{n})] \\
=:& I_1 + I_2 + I_3,
\end{aligned}
$$

where

$$
\begin{aligned}
I_1 =& -\frac{\bar{\eta}_n^\top}{\sqrt{n}}\sum_{i=1}^n [Y_i + g(X_i)] - \sum_{i=1}^n \phi(X_{i-1}, -\frac{\bar{\eta}_n}{\sqrt{n}}) \\
=& -\frac{\bar{\eta}_n^\top}{\sqrt{n}}\sum_{i=1}^n \big[Y_i - \mathbb{E}_\pi[Y_1]\big] - \frac{\bar{\eta}_n^\top}{\sqrt{n}}[g(X_n) - g(X_0)] - \frac{1}{2n}\sum_{i=1}^n \bar{\eta}_n^\top \frac{\partial^2 \phi}{\partial \eta^2}(X_{i-1}, 0)\bar{\eta}_n + o(1) \\
=& -\frac{\bar{\theta}^\top}{\sqrt{n}}\sum_{i=1}^n \big[Y_i - \mathbb{E}_\pi[Y_1]\big] - \frac{\bar{\theta}^\top}{\sqrt{n}}[g(X_n) - g(X_0)] - \frac{1}{2n}\sum_{i=1}^n \bar{\theta}^\top \frac{\partial^2 \phi}{\partial \eta^2}(X_{i-1}, 0)\bar{\theta} + o(1)
\end{aligned}
$$

by the above argument,

$$
I_2 = -\frac{\bar{\theta}_n^\top - \bar{\eta}_n^\top}{\sqrt{n}}\sum_{i=1}^n [Y_i - \frac{\partial \psi}{\partial \theta}(X_{i-1}, X_i, 0)] = (\bar{\theta}_n^\top - \bar{\eta}_n^\top)O(1) = o(1),
$$

and

$$
I_3 = -\frac{1}{2}\sum_{i=1}^n [\frac{1}{n}\bar{\theta}_n^\top \frac{\partial^2 \psi}{\partial \theta^2}(X_{i-1}, X_i, 0)\bar{\theta}_n - \frac{1}{n}\bar{\eta}_n^\top \frac{\partial^2 \psi}{\partial \theta^2}(X_{i-1}, X_i, 0)\bar{\eta}_n + o(\frac{1}{n})] = o(1).
$$

Combining all the above together, we show that asymptotically $LR_n(-\frac{\bar{\theta}_n}{\sqrt{n}}, -\frac{\bar{\eta}_n}{\sqrt{n}})$ converges to the same distribution as $LR_n(-\frac{\bar{\theta}}{\sqrt{n}}, -\frac{\bar{\theta}}{\sqrt{n}})$. The proof is completed. □

Finally, let us prove Theorem 2. Recall that Assumptions EC.1, EC.2, and EC.5 guarantee the existence of ψ and g.

Given the link function (13) with $\theta = -\frac{\bar{\theta}}{\sqrt{n}}$, $\eta = -\frac{\bar{\eta}}{\sqrt{n}}$, the log-likelihood ratio is

$$\begin{aligned}
&LR_n(-\frac{\bar{\theta}}{\sqrt{n}}, -\frac{\bar{\eta}}{\sqrt{n}}) \\
=&-\frac{\bar{\theta}^\top}{\sqrt{n}} S_n + \sum_{i=1}^{n} \big\{ -\psi(X_{i-1}, X_i, -\frac{\bar{\theta}}{\sqrt{n}}) - \frac{\bar{\eta}^\top}{\sqrt{n}} [\frac{\partial \psi}{\partial \theta}(X_{i-1}, X_i, 0) + \Delta(X_i, 0)] - \phi(X_{i-1}, -\frac{\bar{\eta}}{\sqrt{n}}) \big\} \\
=&-\frac{\bar{\theta}^\top}{\sqrt{n}} S_n + \sum_{i=1}^{n} \big\{ \frac{\bar{\theta}^\top - \bar{\eta}^\top}{\sqrt{n}} \frac{\partial \psi}{\partial \theta}(X_{i-1}, X_i, 0) - \frac{\bar{\eta}^\top}{\sqrt{n}} \Delta(X_i, 0) + \frac{\bar{\eta}^\top}{\sqrt{n}} \frac{\partial \phi}{\partial \eta}(X_{i-1}, 0) \\
&+ \frac{1}{2n} [-\bar{\theta}^\top \frac{\partial^2 \psi}{\partial \theta^2}(X_{i-1}, X_i, 0)\bar{\theta} - \bar{\eta}^\top \frac{\partial^2 \phi}{\partial \eta^2}(X_{i-1}, 0)\bar{\eta}] + o(\frac{1}{n}) \big\} \\
=&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big\{ -\bar{\theta}^\top Y_i + (\bar{\theta}^\top - \bar{\eta}^\top) \frac{\partial \psi}{\partial \theta}(X_{i-1}, X_i, 0) - \bar{\eta}^\top \Delta(X_i, 0) + \bar{\eta}^\top \frac{\partial \phi}{\partial \eta}(X_{i-1}, 0) \big\} \\
&- \frac{1}{2n} \sum_{i=1}^{n} [\bar{\theta}^\top \frac{\partial^2 \psi}{\partial \theta^2}(X_{i-1}, X_i, 0)\bar{\theta} + \bar{\eta}^\top \frac{\partial^2 \phi}{\partial \eta^2}(X_{i-1}, 0)\bar{\eta}] + o(1).
\end{aligned}$$

In this case,

$$\mathbb{E}_x\big[e^{\eta^\top [\frac{\partial \psi}{\partial \theta}(X_0, X_1, 0) + \Delta(X_1, 0)]}\big] = e^{\phi(x,\eta)},\ \mathbb{E}[e^{\theta^\top Y_1} | X_0, X_1] = e^{\psi(X_0, X_1, \theta)},$$

hence

$$\mathbb{E}_x[\frac{\partial \psi}{\partial \theta}(X_0, X_1, 0) + \Delta(X_1, 0)] = \frac{\partial \phi}{\partial \eta}(x, 0),\ \mathbb{E}[Y_1 | X_0, X_1] = \frac{\partial \psi}{\partial \theta}(X_0, X_1, 0).$$

$$\mathrm{Var}_x[\frac{\partial \psi}{\partial \theta}(X_0, X_1, 0) + \Delta(X_1, 0)] = \frac{\partial^2 \phi}{\partial \eta^2}(x, 0),\ \mathrm{Var}[Y_1 | X_0, X_1] = \frac{\partial^2 \psi}{\partial \theta^2}(X_0, X_1, 0).$$

Under the stationary distribution π of the underlying Markov chain, then jointly, $\{(X_{i-1}, X_i, Y_i), i \geqslant 0\}$ is also a Markov chain. Following the similar argument in the proof of Lemma EC.4, $LR_n(-\frac{\bar{\theta}}{\sqrt{n}}, -\frac{\bar{\eta}}{\sqrt{n}})$ converges to a normal distribution. Moreover, we can find the asymptotic mean of $LR_n(-\frac{\bar{\theta}}{\sqrt{n}}, -\frac{\bar{\eta}}{\sqrt{n}})$ is $-\frac{1}{2}\mathbb{E}_\pi[\bar{\theta}^\top \frac{\partial^2 \psi}{\partial \theta^2}(X_0, X_1, 0)\bar{\theta} + \bar{\eta}^\top \frac{\partial^2 \phi}{\partial \eta^2}(X_0, 0)\bar{\eta}]$, and the asymptotic variance is $\mathbb{E}_\pi[\bar{\theta}^\top \frac{\partial^2 \psi}{\partial \theta^2}(X_0, X_1, 0)\bar{\theta} + \bar{\eta}^\top \frac{\partial^2 \phi}{\partial \eta^2}(X_0, 0)\bar{\eta}]$. This proves that our tilting family also belongs to LAN family.

In particular, by choosing $\theta = \eta = -\frac{\bar{\theta}}{\sqrt{n}}$, we will claim that $LR_n(-\frac{\bar{\theta}}{\sqrt{n}}, -\frac{\bar{\theta}}{\sqrt{n}}) = \overline{LR}_n(-\frac{\bar{\theta}}{\sqrt{n}}) + o(1)$, where

$$\overline{LR}_n(\theta) = \theta^\top [S_n - n\Lambda'(0) + \sum_{i=1}^{n} \big(\Delta(X_i, 0) - \Delta(X_{i-1}, 0)\big)] - \frac{n}{2}\theta^\top \Lambda''(0)\theta.$$

To see this, consider

$$\begin{aligned}
&LR_n(-\frac{\bar{\theta}}{\sqrt{n}}, -\frac{\bar{\theta}}{\sqrt{n}}) - \overline{LR}_n(-\frac{\bar{\theta}}{\sqrt{n}}) - \frac{1}{2}\bar{\theta}^\top \Lambda''(0)\bar{\theta} \\
=&\sum_{i=1}^{n} \Big[-\frac{\bar{\theta}^\top}{\sqrt{n}} \Delta(X_{i-1}, 0) - \frac{\bar{\theta}^\top}{\sqrt{n}} \Lambda'(0) - \psi(X_{i-1}, X_i, -\frac{\bar{\theta}}{\sqrt{n}}) - \frac{\bar{\theta}^\top}{\sqrt{n}} \frac{\partial \psi}{\partial \theta}(X_{i-1}, X_i, 0) - \phi(X_{i-1}, -\frac{\bar{\theta}}{\sqrt{n}}) \Big] \\
=&\sum_{i=1}^{n} \Big[-\frac{1}{2n} \bar{\theta}^\top [\frac{\partial^2 \psi}{\partial \theta^2}(X_{i-1}, X_i, 0) + \frac{\partial^2 \phi}{\partial \eta^2}(X_{i-1}, 0)]\bar{\theta} + \frac{\bar{\theta}^\top}{\sqrt{n}} [\frac{\partial \phi}{\partial \eta}(X_{i-1}, 0) - \Delta(X_{i-1}, 0) - \Lambda'(0)] \Big].
\end{aligned}$$

By definition of ϕ, we have

$$e^{\phi(x,\eta)} = \mathbb{E}\big[e^{\eta^\top [\frac{\partial \psi}{\partial \theta}(X_0, X_1, 0) + \Delta(X_1, 0)]} | X_0 = x\big],$$

hence $\frac{\partial\phi}{\partial\eta}(x,0)=\mathbb{E}[\frac{\partial\psi}{\partial\theta}(X_0,X_1,0)+\Delta(X_1,0)|X_0=x]$. On the other hand, from (i) we get

$$\Lambda'(0)+\Delta(x,0)=\mathbb{E}[Y_1+\Delta(X_1,0)|X_0=x]=\mathbb{E}[\frac{\partial\psi}{\partial\theta}(X_0,X_1,0)+\Delta(X_1,0)|X_0=x].$$

Therefore, $\Lambda'(0)+\Delta(x,0)=\frac{\partial\phi}{\partial\eta}(x,0)$.

Note that $\overline{LR}_n(-\frac{\bar\theta}{\sqrt{n}})=-\frac{\bar\theta^\top}{\sqrt{n}}[S_n-n\Lambda'(0)]-\frac{\bar\theta^\top}{\sqrt{n}}[\Delta(X_n,0)-\Delta(X_0,0)]+c=-\frac{\bar\theta^\top}{\sqrt{n}}[S_n-n\Lambda'(0)]+c+o(1)$. Therefore, by Lemma EC.3, the proposed tilting probability is asymptotically optimal.

Lastly, we relax the condition to $\theta=-\frac{\bar\theta_n}{\sqrt{n}}$, $\eta=-\frac{\bar\eta_n}{\sqrt{n}}$, and $\lim_{n\to\infty}\bar\theta_n=\lim_{n\to\infty}\bar\eta_n=\bar\theta<\infty$. The rest of the proof follows the same argument as in the proof of Lemma EC.4. □

EC.9.6. Proof of Lemma 3

Under Assumption EC.4 $r(x;\alpha)$ is analytic.

We start from the definition of eigenvalue and eigenfunction in (4). Notice that $\Lambda(0)=0$ and $r(x,0)$ is normalized to be 1. We consider the Taylor expansion of both $r(x,\theta)$ and $\Lambda(\theta)$ at any point θ_0:

$$e^{\Lambda(\theta)}\approx e^{\Lambda(\theta_0)}+e^{\Lambda(\theta_0)}\Lambda'(\theta_0)^\top(\theta-\theta_0)+o(\theta-\theta_0),$$

$$r(x,\theta)\approx r(x,\theta_0)+\frac{\partial r}{\partial\theta}(x,\theta_0)^\top(\theta-\theta_0)+o(\theta-\theta_0).$$

Therefore, under Assumption EC.1, (4) becomes

$$\begin{aligned}&\big[e^{\Lambda(\theta_0)}+e^{\Lambda(\theta_0)}\Lambda'(\theta_0)^\top(\theta-\theta_0)+o(\theta-\theta_0)\big]\big[r(x,\theta_0)+\frac{\partial r}{\partial\theta}(x,\theta_0)^\top(\theta-\theta_0)+o(\theta-\theta_0)\big]\\ =&\mathbb{E}_x\bigg[\big(e^{\theta_0^\top Y_1}+e^{\theta_0^\top Y_1}Y_1^\top(\theta-\theta_0)+o(\theta-\theta_0)\big)\big[r(X_1,\theta_0)+\frac{\partial r}{\partial\theta}(X_1,\theta_0)^\top(\theta-\theta_0)+o(\theta-\theta_0)\big]\bigg].\end{aligned}$$

The leading order terms are canceled. Then comparing $O(\theta-\theta_0)$ term leads to

$$e^{\Lambda(\theta_0)}\frac{\partial r}{\partial\theta}(x,\theta_0)+e^{\Lambda(\theta_0)}\Lambda'(\theta_0)r(x,\theta_0)=\mathbb{E}_x\big[e^{\theta_0^\top Y_1}\frac{\partial r}{\partial\theta}(X_1,\theta_0)+e^{\theta_0 Y_1}Y_1r(X_1,\theta_0)\big].$$

Denote $\Delta(x,\theta)=\frac{\partial r}{\partial\theta}(x,\theta)$, then the above equality becomes

$$\Delta(x,\theta_0)+\Lambda'(\theta_0)=\mathbb{E}_x\big[e^{\theta_0^\top Y_1-\Lambda(\theta_0)}\frac{r(X_1,\theta_0)}{r(x,\theta_0)}\big(\Delta(X_1,\theta_0)+Y_1\big)\big].$$

Recall the definition of the probability $\mathbb{P}_\theta$ in (4), we obtain

$$\mathbb{E}_x^{\mathbb{P}_{\theta_0}}[\Delta(X_1,\theta_0)-\Delta(x,\theta_0)+Y_1]=\Lambda'(\theta_0).$$

Moreover, since the right hand side does not depend on the initial state x, we have taken the initial state as any arbitrary distribution ν. Then

$$\mathbb{E}_\nu^{\mathbb{P}_{\theta_0}}[\Delta(X_1,\theta_0)-\Delta(x,\theta_0)+Y_1]=\Lambda'(\theta_0).$$

In particular, if we take the initial distribution to be the stationary distribution under $\mathbb{P}_\theta$, denoted by π_θ, then $\mathbb{E}_{\pi_\theta}^{\mathbb{P}_{\theta_0}}[\Delta(X_1,\theta_0)-\Delta(x,\theta_0)]=0$, hence $\mathbb{E}_{\pi_\theta}^{\mathbb{P}_\theta}[Y_1]=\Lambda'(\theta)$.

In particular, take $\theta_0=0$ to obtain

$$\mathbb{E}_x[\Delta(X_1,0)-\Delta(x,0)]=\Lambda'(0)-\mathbb{E}_x[Y_1]=\mathbb{E}_\pi[Y_1]-\mathbb{E}_x[Y_1],$$

where π is the stationary distribution of the underlying Markov chain $\{X_n, n\geqslant 0\}$. The left hand side can be represented as $(\mathcal{P}_0-I)\Delta(x,0)$, where I is the identity operator and $\mathcal{P}_0$ is the operator defined in (3). □

EC.9.7. Proof of Theorem 3

(i) We first consider the fixed-time event.

Under the asymptotic normal regime, by the Borjesson-Sundberg bound on the Mills ratio, we have

$$\Phi(-\ell)\geqslant \frac{\ell}{1+\ell^2}\varphi(\ell)=\frac{\ell}{(1+\ell^2)\sqrt{2\pi\sigma^2}}e^{-\frac{\ell^2}{2}}.$$

To prove the bounded relative or logarithmic efficiency, it suffices to find suitable upper bound for the variance of the estimator.

When $S_n=\sum_{k=1}^n Y_k$, and $\{(X_n,S_n), n\geqslant 0\}$ forms a Markov random walk as described in Section 2. The classical exponential tilting family is given by (5), with $r(\cdot,\theta)$ and $\Lambda(\theta)$ eigenfunction and eigenvalue respectively. Note that under Assumptions EC.1 and EC.4, $\Lambda(\theta)$ is analytic for $\theta\in\Theta$.

Without loss of generality, we assume $\ell>0$, and $\sqrt{n}\frac{S_n}{n}\to\mathcal{N}(0,1)$ in the later development, and restrict to the one-dimensional case $S_n\in\mathbb{R}$. Here we assume the stationary mean of S_1 is 0 and the stationary variance of S_1 is 1. Under Assumptions EC.1 to EC.4, by Theorem 17.0.1 of Meyn and Tweedie (2012), we have $\frac{S_n}{\sqrt{n}}\to\mathcal{N}(0,1)$.

By Theorem 4.1 of Ney and Nummelin (1987), we have the relation between the eigenvalue and the mean and variance of the limiting distribution as

$$0=\mathbb{E}_\pi[Y_1]=\Lambda'(0),\ 1=\Lambda''(0).$$

And we still can define the Legendre transformation of Λ by

$$\hat{\Lambda}(c)=\sup_{\theta\geqslant 0}\{c\theta-\Lambda(\theta)\}=c\Lambda'^{-1}(c)-\Lambda\left(\Lambda'^{-1}(c)\right).$$

Recall that the importance sampling estimator in the classical exponential tilting, $\hat{Z}(n,\ell;\theta)$, is to sample under $\mathbb{P}_\theta$, with $\theta>0$ and is given by $\mathbb{1}_{\{\sqrt{n}\frac{S_n}{n}>\ell\}}\frac{r(X_0,\theta)}{r(X_n,\theta)}e^{-\theta S_n+n\Lambda(\theta)}$. It has variance

$$\begin{aligned}\mathrm{Var}\left(\hat{Z}(n,\ell;\theta)\right)=&\mathbb{E}_\nu^{\mathbb{P}_\theta}\left[\mathbb{1}_{\{\sqrt{n}\frac{S_n}{n}>\ell\}}e^{-2\theta S_n+2n\Lambda(\theta)+2\log r(X_0,\theta)-2\log r(X_n,\theta)}\right]-\left(z(n,x)\right)^2\\ \leqslant&\mathbb{E}_\nu^{\mathbb{P}_\theta}\left[\mathbb{1}_{\{\sqrt{n}\frac{S_n}{n}>\ell\}}e^{-2\theta S_n+2n\Lambda(\theta)+2\log r(X_0,\theta)-2\log r(X_n,\theta)}\right]\\ \leqslant&\exp\left\{-2n\left[\frac{\ell}{\sqrt{n}}\theta-\Lambda(\theta)\right]\right\}\mathbb{E}_\nu^{\mathbb{P}_\theta}\left[\left(\frac{r(X_0,\theta)}{r(X_n,\theta)}\right)^2\right].\end{aligned}$$

By picking $\theta^*_{n,\ell} = \Lambda'^{-1}(\frac{\ell}{\sqrt{n}})$, we have

$$\mathrm{Var}\left(\hat{Z}(n,\ell;\theta^*_{n,\ell})\right) \leqslant \exp\left\{-2n\hat{\Lambda}(\frac{\ell}{\sqrt{n}})\right\} \mathbb{E}_\nu^{\mathbb{P}_{\theta^*_{n,\ell}}}\left[\left(\frac{r(X_0,\theta^*_{n,\ell})}{r(X_n,\theta^*_{n,\ell})}\right)^2\right].$$

Moreover, by Taylor's expansion, we have

$$\theta^*_{n,\ell} = \left[\frac{1}{\Lambda''\left(\Lambda'^{-1}(0)\right)}\frac{\ell}{\sqrt{n}} + o(\frac{\ell}{\sqrt{n}})\right] = \frac{\ell}{\sqrt{n}} + o(\frac{\ell}{\sqrt{n}}) \to 0.$$

Hence

$$-2n\hat{\Lambda}(\frac{\ell}{\sqrt{n}}) = -2n\left[\hat{\Lambda}(0) + \hat{\Lambda}'(0)\frac{\ell}{\sqrt{n}} + \frac{1}{2}\hat{\Lambda}''(0)\frac{\ell^2}{n} + o(\frac{\ell^2}{n})\right].$$

In addition, Assumption EC.6 implies $\mathbb{E}_\nu^{\mathbb{P}_{\theta^*_{n,\ell}}}\left[\left(\frac{r(X_0,\theta^*_{n,\ell})}{r(X_n,\theta^*_{n,\ell})}\right)^2\right] \leqslant C$ for some constant C as $\theta^*_{n,\ell} \to 0$.

Note that as $\ell \to \infty, n \to \infty, \ell/\sqrt{n} \to 0$, $\theta^*_{n,\ell} \to 0$. Therefore,

$$\begin{aligned}
&\limsup_{\ell\to\infty,n\to\infty,\ell/\sqrt{n}\to 0} \frac{\mathrm{Var}\left(\hat{Z}(n,\ell;\theta^*_{n,\ell})\right)}{\left(\Phi(-\ell)\right)^{2-\epsilon}} \\
\leqslant &\limsup_{\ell\to\infty,n\to\infty,\ell/\sqrt{n}\to 0} C\left(\frac{(1+\ell^2)\sqrt{2\pi}}{\ell}\right)^{2-\epsilon} e^{-2n[\hat{\Lambda}(\frac{\ell}{\sqrt{n}})-\frac{1}{2}\frac{\ell^2}{n}+\frac{\epsilon}{4}\frac{\ell^2}{n}]} \\
= &\limsup_{\ell\to\infty,n\to\infty,\ell/\sqrt{n}\to 0} C\left(\frac{(1+\ell^2)\sqrt{2\pi}}{\ell}\right)^{2-\epsilon} e^{-2n[o(\frac{\ell^2}{n})+\frac{\epsilon}{4}\frac{\ell^2}{n}]} = 0,
\end{aligned}$$

where the last limit equals to zero because for sufficiently large n, ℓ, we have $n[o(\frac{\ell^2}{n}) + \frac{\epsilon}{4}\frac{\ell^2}{n}] \geqslant \frac{\epsilon}{8}\ell^2$. Hence the sequence of estimator $\{\hat{Z}(n,\ell;\theta^*_{n,\ell})\}$ has logarithmic efficiency.

Next, we show the sequence of estimator $\{\hat{Z}(n,\ell;\theta^*_{n,\ell})\}$ belongs to the LAN family (based on Definition 1). The log-likelihood ratio is

$$\begin{aligned}
&\theta^*_{n,\ell}S_n - n\Lambda(\theta^*_{n,\ell}) + \log r(X_n,\theta^*_{n,\ell}) - \log r(X_0,\theta^*_{n,\ell}) \\
= &\Lambda'^{-1}(\frac{\ell}{\sqrt{n}})S_n - n\Lambda\left(\Lambda'^{-1}(\frac{\ell}{\sqrt{n}})\right) + \log r\left(X_n,\Lambda'^{-1}(\frac{\ell}{\sqrt{n}})\right) - \log r\left(X_0,\Lambda'^{-1}(\frac{\ell}{\sqrt{n}})\right) \\
= &\left[\frac{1}{\Lambda''\left(\Lambda'^{-1}(0)\right)}\frac{\ell}{\sqrt{n}} + o(\frac{1}{\sqrt{n}})\right]S_n - n\left[\frac{1}{2}\Lambda''(0)\left[\frac{1}{\Lambda''\left(\Lambda'^{-1}(0)\right)}\frac{\ell}{\sqrt{n}} + o(\frac{1}{\sqrt{n}})\right]^2 + o(\frac{1}{n})\right] \\
&+ \left[\frac{\partial r}{\partial\theta}(X_n,0) - \frac{\partial r}{\partial\theta}(X_0,0)\right]\left[\frac{1}{\Lambda''\left(\Lambda'^{-1}(0)\right)}\frac{\ell}{\sqrt{n}} + o(\frac{1}{\sqrt{n}})\right] \\
= &\ell\frac{1}{\sqrt{n}}S_n + O(\frac{1}{\sqrt{n}}) - \frac{\ell^2}{2} + o(1) \to \mathcal{N}(-\frac{\ell^2}{2},\ell^2) \text{ as } n\to\infty. \qquad \text{(EC.12)}
\end{aligned}$$

Note that under Assumptions EC.1 to EC.4, the asymptotic normality in (EC.12) comes from Theorem 17.0.1 of Meyn and Tweedie (2012).

Now suppose there is another sequence of estimator $\{\hat{Z}_{opt}(n,\ell)\}$ belonging to the LAN family and achieving the smallest asymptotic variance. Then it means

$$\limsup_{\ell\to\infty,n\to\infty,\ell/\sqrt{n}\to 0}\frac{\mathrm{Var}\left(\hat{Z}_{opt}(n,\ell)\right)}{\left(\Phi(-\ell)\right)^{2-\epsilon}}\leqslant\limsup_{\ell\to\infty,n\to\infty,\ell/\sqrt{n}\to 0}\frac{\mathrm{Var}\left(\hat{Z}(n,\ell;\theta^*_{n,\ell})\right)}{\left(\Phi(-\ell)\right)^{2-\epsilon}}.$$

Hence $\{\hat{Z}_{opt}(n,\ell)\}$ also has logarithmic efficiency.

(ii) Next we consider the first-passage time event.

When $S_n=\sum_{k=1}^n Y_k$, and $\{(X_n,S_n),n\geqslant 0\}$ forms a Markov random walk as described in Section 2. Recall that the importance sampling estimator in the classical exponential tilting, $\hat{Z}(b,\ell;\theta)$, is to sample under $\mathbb{P}_\theta$, with $\theta>0$ and is given by $\mathbb{1}_{\{\tau_b<\sqrt{\frac{b\sigma^2}{\mu^3}}\ell+\frac{b}{\mu}\}}\frac{r(X_0,\theta)}{r(X_{\tau_b},\theta)}e^{-\theta S_{\tau_b}+\tau_b\Lambda(\theta)}$, where $\theta\geqslant 0$. Since $\Lambda'(0)=\mu>0$, $\Lambda''(0)=\sigma^2<\infty$, we know when $\theta>0$ is sufficiently close to 0, $\Lambda(\theta)>0$.

In this case, the importance sampling estimator in the classical exponential tilting has variance

$$\begin{aligned}\mathrm{Var}\left(\hat{Z}(b,\ell;\theta)\right)=&\mathbb{E}_\nu^{\mathbb{P}_\theta}[\mathbb{1}_{\{\tau_b<\sqrt{\frac{b\sigma^2}{\mu^3}}\ell+\frac{b}{\mu}\}}e^{-2\theta S_{\tau_b}+2\tau_b\Lambda(\theta)+2\log r(X_0,\theta)-2\log r(X_{\tau_b},\theta)}]-\left(z(n,b)\right)^2\\ \leqslant&\mathbb{E}_\nu^{\mathbb{P}_\theta}[\mathbb{1}_{\{\tau_b<\sqrt{\frac{b\sigma^2}{\mu^3}}\ell+\frac{b}{\mu}\}}e^{-2\theta b+2[\sqrt{\frac{b\sigma^2}{\mu^3}}\ell+\frac{b}{\mu}]\Lambda(\theta)+2\log r(X_0,\theta)-2\log r(X_{\tau_b},\theta)}]\\ \leqslant&\exp\left\{-2\left(\sqrt{\frac{b\sigma^2}{\mu^3}}\ell+\frac{b}{\mu}\right)\left[\frac{b}{\sqrt{\frac{b\sigma^2}{\mu^3}}\ell+\frac{b}{\mu}}\theta-\Lambda(\theta)\right]\right\}\mathbb{E}_\nu^{\mathbb{P}_\theta}\left[\left(\frac{r(X_0,\theta)}{r(X_{\tau_b},\theta)}\right)^2\right].\end{aligned}$$

By picking $\theta^*_{b,\ell}=\Lambda'^{-1}(\frac{b}{\sqrt{\frac{b\sigma^2}{\mu^3}}\ell+\frac{b}{\mu}})=\Lambda'^{-1}(\frac{\mu}{\sigma\ell/\sqrt{b\mu}+1})=\frac{1}{\sigma^2}\left(-\frac{\sqrt{\mu}\sigma}{\sqrt{b}}\ell+\frac{\sigma^2}{2b}\ell+o(\frac{1}{b})\right)+O(\frac{\ell}{b})$, we have

$$\mathrm{Var}\left(\hat{Z}(b,\ell;\theta^*_{b,\ell})\right)\leqslant\exp\left\{-2\left(\sqrt{\frac{b\sigma^2}{\mu^3}}\ell+\frac{b}{\mu}\right)\hat{\Lambda}\left(\frac{\mu}{\sigma\ell/\sqrt{b\mu}+1}\right)\right\}\mathbb{E}_\nu^{\mathbb{P}_{\theta^*_{b,\ell}}}\left[\left(\frac{r(X_0,\theta^*_{b,\ell})}{r(X_{\tau_b},\theta^*_{b,\ell})}\right)^2\right].$$

Moreover, by Taylor's expansion,

$$\begin{aligned}\hat{\Lambda}(\frac{\mu}{\sigma\ell/\sqrt{b\mu}+1})=&\Lambda'^{-1}(\frac{\mu}{\sigma\ell/\sqrt{b\mu}+1})\frac{\mu}{\sigma\ell/\sqrt{b\mu}+1}-\Lambda\left(\Lambda'^{-1}(\frac{\mu}{\sigma\ell/\sqrt{b\mu}+1})\right)\\ =&\Lambda'^{-1}(\frac{\mu}{\sigma\ell/\sqrt{b\mu}+1})\left[\mu-\frac{\sqrt{\mu}\sigma}{\sqrt{b}}\ell+\frac{\sigma^2}{2b}\ell+o(\frac{1}{b})\right]\\ &-\mu\Lambda'^{-1}(\frac{\mu}{\sigma\ell/\sqrt{b\mu}+1})-\frac{\sigma^2}{2}\left(\Lambda'^{-1}(\frac{\mu}{\sigma\ell/\sqrt{b\mu}+1})\right)^2\\ =&\left[\frac{1}{\sigma^2}\left(-\frac{\sqrt{\mu}\sigma}{\sqrt{b}}\ell+\frac{\sigma^2}{2b}\ell+o(\frac{1}{b})\right)+O(\frac{1}{b})\right]\left[-\frac{\sqrt{\mu}\sigma}{\sqrt{b}}\ell+\frac{\sigma^2}{2b}\ell+o(\frac{1}{b})\right]\\ &-\frac{\sigma^2}{2}\left[\frac{1}{\sigma^2}\left(-\frac{\sqrt{\mu}\sigma}{\sqrt{b}}\ell+\frac{\sigma^2}{2b}\ell+o(\frac{1}{b})\right)+O(\frac{1}{b})\right]^2\\ =&\frac{\mu}{2b}\ell^2+o(\frac{\ell^2}{b}).\end{aligned}$$

Hence

$$-2\left(\sqrt{\frac{b\sigma^2}{\mu^3}}\ell+\frac{b}{\mu}\right)\hat{\Lambda}(\frac{\mu}{\sigma\ell/\sqrt{b\mu}+1})=-2\left(\sqrt{\frac{b\sigma^2}{\mu^3}}\ell+\frac{b}{\mu}\right)\left(\frac{\mu}{2b}\ell^2+o(\frac{1}{b})\right)=-\ell^2+O(\frac{\ell}{\sqrt{b}}).$$

In addition, Assumption EC.6 implies $\mathbb{E}_\nu^{\mathbb{P}_{\theta^*_{b,\ell}}}\left[\left(\frac{r(X_0,\theta^*_{b,\ell})}{r(X_{\tau_b},\theta^*_{b,\ell})}\right)^2\right] \leqslant C$ for some constant C as $\theta^*_{b,\ell}\to 0$.

Note that as $\ell\to-\infty, b\to\infty, \ell/\sqrt{b}\to 0$, $\theta^*_{b,\ell}\to 0$. Therefore,

$$\begin{aligned}
&\limsup_{\ell\to-\infty,n\to\infty,\ell/\sqrt{n}\to 0}\frac{\mathrm{Var}\left(\hat{Z}(n,\ell;\theta^*_{b,\ell})\right)}{\left(\Phi(\ell)\right)^{2-\epsilon}}\\
\leqslant&\limsup_{\ell\to-\infty,n\to\infty,\ell/\sqrt{n}\to 0} C\left(\sqrt{2\pi\ell}\right)^{2-\epsilon} e^{-2n[\hat{\Lambda}(\frac{\ell}{\sqrt{n}})-\frac{1}{2}\frac{\ell^2}{n}+\frac{\epsilon}{4}\frac{\ell^2}{n}]}\\
=&\limsup_{\ell\to-\infty,n\to\infty,\ell/\sqrt{n}\to 0} C\left(\sqrt{2\pi\ell}\right)^{2-\epsilon} e^{-2n[o(\frac{\ell^2}{n})+\frac{\epsilon}{4}\frac{\ell^2}{n}]}=0,
\end{aligned}$$

for the same reason as in the fixed-time event case. Hence the sequence of estimator $\{\hat{Z}(b,\ell;\theta^*_{b,\ell})\}$ has logarithmic efficiency.

Next, we show the sequence of estimator $\{\hat{Z}(b,\ell;\theta^*_{b,\ell})\}$ belongs to the LAN family (based on Definition 1 (ii)). Denote $R(b)=S_{\tau_b}-b$ as the overshoot. The log-likelihood ratio is

$$\begin{aligned}
&\theta^*_{b,\ell}S_{\tau_b}-\tau_b\Lambda(\theta^*_{b,\ell})+\log r(X_{\tau_b},\theta^*_{b,\ell})-\log r(X_0,\theta^*_{b,\ell})\\
=&\ \Lambda'^{-1}(\frac{\mu}{\sigma\ell/\sqrt{b\mu}+1})[b+R(b)]-\tau_b\Lambda\left(\Lambda'^{-1}(\frac{\mu}{\sigma\ell/\sqrt{b\mu}+1})\right)\\
&+\left[\frac{\partial r}{\partial\theta}(X_{\tau_b},0)-\frac{\partial r}{\partial\theta}(X_0,0)\right]\left[\frac{1}{\sigma^2}\left(-\frac{\sqrt{\mu}\sigma}{\sqrt{b}}\ell+\frac{\sigma^2}{2b}\ell+o(\frac{1}{b})\right)+O(\frac{1}{b})\right]\\
=&\ \Lambda'^{-1}(\frac{\mu}{\sigma\ell/\sqrt{b\mu}+1})[b+R(b)]-\left[\frac{\tau_b-b/\mu}{\sqrt{b\sigma^2/\mu^3}}\sqrt{\frac{b\sigma^2}{\mu^3}}+\frac{b}{\mu}\right]\left[\mu\Lambda'^{-1}(\frac{\mu}{\sigma\ell/\sqrt{b\mu}+1})\right.\\
&\left.+\frac{\sigma^2}{2}\left(\Lambda'^{-1}(\frac{\mu}{\sigma\ell/\sqrt{b\mu}+1})\right)^2+o(\frac{1}{b})\right]+\left[\frac{\partial r}{\partial\theta}(X_{\tau_b},0)-\frac{\partial r}{\partial\theta}(X_0,0)\right]\left[-\frac{\sqrt{\mu}}{\sigma\sqrt{b}}\ell+O(\frac{1}{b})\right]\\
=&\left[-\frac{\sqrt{\mu}}{\sigma\sqrt{b}}\ell+O(\frac{1}{b})\right]R(b)-\frac{b\sigma^2}{2\mu}\left(\frac{\mu}{b\sigma^2}\ell^2+o(\frac{1}{b})\right)-\frac{\tau_b-b/\mu}{\sqrt{b\sigma^2/\mu^3}}\sqrt{\frac{b\sigma^2}{\mu^3}}\left(\frac{\mu}{\sigma^2}(-\frac{\sqrt{\mu}\sigma}{\sqrt{b}}\ell)+O(\frac{1}{b})\right)\\
&+\left[\frac{\partial r}{\partial\theta}(X_{\tau_b},0)-\frac{\partial r}{\partial\theta}(X_0,0)\right]\left[-\frac{\sqrt{\mu}}{\sigma\sqrt{b}}\ell+O(\frac{1}{b})\right]\\
=&\ R(b)O(\frac{1}{b})-\frac{1}{2}\ell^2+\frac{\tau_b-b/\mu}{\sqrt{b\sigma^2/\mu^3}}\ell+O(\frac{1}{b}).
\end{aligned}\tag{EC.13}$$

To handle the asymptotic behavior in (EC.13), we first have that under Assumptions EC.1 to EC.4, Theorems 2 and 3 of Fuh (2004) show that $\{\tau<n, S_n<\gamma n^{1/2}\}$, for a suitable chosen γ, satisfies the asymptotic normality. Note that $\mathbb{P}_\pi(\tau<n)=\mathbb{P}_\pi(S_n\geqslant b)+\mathbb{P}_\pi(\tau<n,S_n<b)$. By using the central limit theorem of $\mathbb{P}_\pi(S_n\geqslant b)$ for a suitable chosen b, we have $\frac{\tau_b-b/\mu}{\sqrt{b\sigma^2/\mu^3}}\to\mathcal{N}(0,1)$ as $b\to\infty$. Next, we consider the overshoot $R(b)$. Under Assumptions EC.1 to EC.4, by Lemma 2 of Fuh (2004), we have $\sup_{b\geqslant 0}E_\pi R(b)\leqslant E_\pi(S_1^+)^2/E_\pi S_1$, thus $R(b)O(\frac{1}{b})\to 0$ in probability as $b\to\infty$. Therefore, $\theta^*_{b,\ell}S_{\tau_b}-\tau_b\psi(\theta^*_{b,\ell})\to\mathcal{N}(-\frac{1}{2}\ell^2,\ell^2)$.

Now suppose there is another sequence of estimator $\{\hat{Z}_{opt}(b,\ell)\}$ belonging to the LAN family and achieving the smallest asymptotic variance. Then it means

$$\limsup_{\ell\to-\infty,n\to\infty,\ell/\sqrt{n}\to 0}\frac{\mathrm{Var}\left(\hat{Z}_{opt}(b,\ell)\right)}{\left(\Phi(\ell)\right)^{2-\epsilon}}\leqslant\limsup_{\ell\to-\infty,n\to\infty,\ell/\sqrt{n}\to 0}\frac{\mathrm{Var}\left(\hat{Z}(n,\ell;\theta^*_{b,\ell})\right)}{\left(\Phi(\ell)\right)^{2-\epsilon}}.$$

Hence $\{\hat{Z}_{opt}(b,\ell)\}$ also has logarithmic efficiency. □

EC.9.8. Proof of Theorem EC.1

We only illustrate the fixed-time event here, and the proof for the first-passage-time event is similar.

From the proof of Theorem 3, we have

$$\mathrm{Var}\left(\hat{Z}(n,\ell;\theta^*_{n,\ell})\right)\leqslant\exp\left\{-2n\hat{\Lambda}(\frac{\ell}{\sqrt{n}})\right\}\mathbb{E}_\nu^{\mathbb{P}_{\theta^*_{n,\ell}}}\left[\left(\frac{r(X_0,\theta^*_{n,\ell})}{r(X_n,\theta^*_{n,\ell})}\right)^2\right],$$

where $\theta^*_{n,\ell}=\Lambda'^{-1}(\frac{\ell}{\sqrt{n}})$. Under Assumption EC.7, by Chan and Lai (2011, Example 1), the eigenfunction $r(x,\theta)$ is uniformly positive and bounded for all $\theta\in\Theta$. Hence, $\mathrm{Var}\left(\hat{Z}(n,\ell;\theta^*_{n,\ell})\right)\leqslant C\exp\left\{-2n\hat{\Lambda}(\frac{\ell}{\sqrt{n}})\right\}$.

On the other hand, let us examine a lower bound for $z(n,\ell)$. Note that

$$\begin{aligned}z(n,\ell)=&\mathbb{E}_\nu^{\mathbb{P}_{\theta^*_{n,\ell}}}[\mathbb{1}_{\{\sqrt{n}\frac{S_n}{n}>\ell\}}e^{-\theta^*_{n,\ell}S_n+n\Lambda(\theta^*_{n,\ell})}\frac{r(X_0,\theta^*_{n,\ell})}{r(X_n,\theta^*_{n,\ell})}]\\ \geqslant&C'e^{n\Lambda(\theta^*_{n,\ell})-\theta^*_{n,\ell}\sqrt{n}\ell}\mathbb{E}_\nu^{\mathbb{P}_{\theta^*_{n,\ell}}}[\mathbb{1}_{\{1>\frac{S_n-\sqrt{n}\ell}{\sqrt{n}}>0\}}e^{-\theta^*_{n,\ell}(S_n-\sqrt{n}\ell)}]\\ \geqslant&C'e^{n\Lambda(\theta^*_{n,\ell})-\theta^*_{n,\ell}\sqrt{n}\ell-\theta^*_{n,\ell}\sqrt{n}}=C'\exp\{-n\hat{\Lambda}(\theta^*_{n,\ell})-\theta^*_{n,\ell}\sqrt{n}\}.\end{aligned}$$

Recall that $\theta^*_{n,\ell}=\Lambda'^{-1}(\frac{\ell}{\sqrt{n}})\to\Lambda'^{-1}(c)$, therefore,

$$\begin{aligned}&\limsup_{\ell\to\infty,n\to\infty,\ell/\sqrt{n}\to c}\frac{\mathrm{Var}\left(\hat{Z}(n,\ell)\right)}{\left(z(n,\ell)\right)^{2-\varepsilon}}\\ \leqslant&\limsup_{\ell\to\infty,n\to\infty,\ell/\sqrt{n}\to c}\frac{C}{C'^2}\exp\{-\epsilon n[\hat{\Lambda}(\theta^*_{n,\ell})-\frac{2-\epsilon}{\epsilon}\frac{1}{\sqrt{n}}\theta^*_{n,\ell}]\}=0.\end{aligned}$$

From the proof of Theorem 3, $\hat{Z}(n,\ell;\theta^*_{n,\ell})$ belongs to the same parametric family as $\{\hat{Z}_{opt}(n,\ell)\}$, which minimizes the variance for every (n,ℓ). Then it means

$$\limsup_{\ell\to\infty,n\to\infty,\ell/\sqrt{n}\to 0}\frac{\mathrm{Var}\left(\hat{Z}_{opt}(n,\ell)\right)}{\left(z(n,\ell)\right)^{2-\epsilon}}\leqslant\limsup_{\ell\to\infty,n\to\infty,\ell/\sqrt{n}\to 0}\frac{\mathrm{Var}\left(\hat{Z}(n,\ell;\theta^*_{n,\ell})\right)}{\left(z(n,\ell)\right)^{2-\epsilon}}.$$

Hence $\{\hat{Z}_{opt}(n,\ell)\}$ also has logarithmic efficiency.

References

Asmussen S, Glynn PW (2007) *Stochastic Simulation: Algorithms and Analysis*, volume 57 (Springer Science & Business Media).

Chan HP, Lai TL (2003) Saddlepoint approximations and nonlinear boundary crossing probabilities of Markov random walks. *The Annals of Applied Probability* 13(2):395–429.

Chan HP, Lai TL (2011) A sequential Monte Carlo approach to computing tail probabilities in stochastic models. *The Annals of Applied Probability* 21(6):2315–2342.

Collamore JF (2002) Importance sampling techniques for the multidimensional ruin problem for general Markov additive sequences of random vectors. *The Annals of Applied Probability* 12(1):382–421.

De Boer PT, Kroese DP, Mannor S, Rubinstein RY (2005) A tutorial on the cross-entropy method. *Annals of Operations Research* 134:19–67.

Djehiche B, Hult H, Nyquist P (2022) Importance sampling for a simple Markovian intensity model using subsolutions. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 32(2):1–25.

Dupuis P, Spiliopoulos K, Wang H (2012) Importance sampling for multiscale diffusions. *Multiscale Modeling & Simulation* 10(1):1–27.

Fouque JP, Tullie T (2002) Variance reduction for Monte Carlo simulation in a stochastic volatility environment. *Quantitative Finance* 2(1):24–30.

Fuh CD (2004) Uniform Markov renewal theory and ruin probabilities in Markov random walks. *The Annals of Applied Probability* 14(3):1202–1241.

Fuh CD, Hu I (2004) Efficient importance sampling for events of moderate deviations with applications. *Biometrika* 91(2):471–490.

Fuh CD, Hu I (2007) Estimation in hidden Markov models via efficient importance sampling. *Bernoulli* 13(2):492–513.

Jin P, Mandrekar V, Rüdiger B, Trabelsi C (2013) Positive Harris recurrence of the CIR process and its applications. *Communications on Stochastic Analysis* 7(3):4.

Le A, Singleton KJ, Dai Q (2010) Discrete-time affine^Q term structure models with generalized market prices of risk. *The Review of Financial Studies* 23(5):2184–2227.

Meitz M, Saikkonen P (2008) Stability of nonlinear AR-GARCH models. *Journal of Time Series Analysis* 29(3):453–475.

Meyn SP, Tweedie RL (2012) *Markov Chains and Stochastic Stability* (Springer Science & Business Media).

Ney P, Nummelin E (1987) Markov additive processes I. Eigenvalue properties and limit theorems. *The Annals of Probability* 15(2):561–592.

Rubinstein RY, Kroese DP (2004) *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning* (Springer Science & Business Media).