A SEQUENTIAL MONTE CARLO APPROACH TO COMPUTING TAIL PROBABILITIES IN STOCHASTIC MODELS

By Hock Peng Chan[*]

National University of Singapore

AND

By Tze Leung Lai[†]

Stanford University

Sequential Monte Carlo methods which involve sequential importance sampling and resampling are shown to provide a versatile approach to computing probabilities of rare events. By making use of martingale representations of the sequential Monte Carlo estimators, we show how resampling weights can be chosen to yield logarithmically efficient Monte Carlo estimates of large deviation probabilities for multidimensional Markov random walks.

1. Introduction. In complex stochastic models, it is often difficult to evaluate probabilities of events of interest analytically and Monte Carlo methods provide a practical alternative. When an event A occurs with a small probability (e.g. 10^{-4}), generating 100 events would require a very large number of events (e.g. 1 million) for direct Monte Carlo computation of $P(A)$. To circumvent this difficulty, one can use importance sampling instead of direct Monte Carlo, changing the measure P to Q under which A is no longer a rare event and evaluating $P(A) = E_Q(L\mathbf{1}_A)$ by $m^{-1}\sum_{i=1}^{m} L_i \mathbf{1}_{A_i}$, where $(L_1, \mathbf{1}_{A_1}), \ldots, (L_m, \mathbf{1}_{A_m})$ are m independent samples drawn from the distribution Q, with L_i being a realization of the likelihood ratio statistic $L := dP/dQ$, which is the importance weight. While large deviations theory has provided important clues for the choice of Q for Monte Carlo evaluation of exceedance probabilities, it has also been demonstrated that importance sampling measures that are consistent with large deviations can perform much worse than direct Monte Carlo; see Glasserman and Wang [18]. Chan and Lai [8] have recently resolved this problem by showing that certain

---

1

mixtures of exponentially twisted measures are asymptotically optimal for importance sampling. For complex stochastic models, however, there are implementation difficulties in using these asymptotically optimal importance sampling measures. Herein we introduce a *sequential importance sampling and resampling* (SISR) procedure to attain a weaker form of asymptotic optimality, namely, logarithmic efficiency; the definitions of *asymptotic optimality* and *logarithmic efficiency* are given in Section 3.

Instead of applying directly the asymptotically optimal importance sampling measure Q that is difficult to sample from, SISR generates m sequential samples from a more tractable importance sampling measure $\widetilde{Q}$ and resamples at every stage t the m sequential sample paths, yielding a modified sample path after resampling. The objective is to approximate the target measure Q by the weighted empirical measure defined by the resampling weights. Details are given in Section 2 for general resampling weights (not necessarily those associated with the asymptotically optimal resampling measure). Section 4 illustrates the SISR method for Monte Carlo computation of exceedance probabilities in a variety of applications, which include boundary crossing probabilities of generalized likelihood ratio statistics and tail probabilities of Markov random walks. These applications demonstrate the versatility of the SISR method and the relative ease of its implementation.

Our SISR procedure to compute probabilities of rare events is closely related to (a) the *interacting particle systems* (IPS) approach introduced by Del Moral and Garnier [14] to compute tail probabilities of the form $P\{V(X_t) \geq a\}$ for a possibly non-homogeneous Markov chain $\{X_t\}$, and (b) the *dynamic importance sampling* method introduced by Dupuis and Wang [16, 17] to compute $P\{S_n/n \in A\}$, where $S_n = \sum_{t=1}^{n} g(X_t)$ and $\{X_n\}$ is a uniformly recurrent Markov chain with stationary distribution π such that $\int g(x)d\pi(x) \notin A$. Both IPS and dynamic importance sampling generate the X_i sequentially. Dynamic importance sampling uses an adaptive change of measures, based on the simulated paths up to each time $t \leq n$. A recent method closely related to dynamic importance sampling is sequential state-dependent change of measures, introduced by Blanchet and Glynn [3] for Monte Carlo evaluation of tail probabilities of the maximum of heavy-tailed random walks. The IPS approach uses "mutation" to sample $\widetilde{X}_{t+1}^{(i)}$ (conditional on the $X_1^{(i)}, \ldots, X_t^{(i)}$ already generated) from the original measure P and then uses "selection" to draw m i.i.d. samples from $\{(X_1^{(i)}, \ldots, X_t^{(i)}, \widetilde{X}_{t+1}^{(i)}) : 1 \leq i \leq m\}$ according to a Boltzmann-Gibbs particle measure. The theory of IPS in [14] focuses on tail probabilities of $V(X_t)$ for fixed t as described in Section 2 rather than large deviation probabilities

of $g(S_n/n)$ for large n as considered in Section 3. Our SISR procedure is motivated by rare events of the general form $\{\mathbf{X}_n \in \Gamma\}$ that involves the entire sample path $\mathbf{X}_n = (X_1, \ldots, X_n)$ and includes $\{V(X_n) \geq a\}$ and $\{S_n/n \in A\}$ considered by Del Moral and Garnier, Dupuis and Wang as special cases. The sequential importance sampling component of SISR uses an easily implementable approximation $\widetilde{Q}$ of Q; in many cases it simply uses $\widetilde{Q} = P$. Thus, it is quite different from dynamic importance sampling, even though both yield logarithmically efficient Monte Carlo estimates of $P\{S_n/n \in A\}$.

2. Sequential importance sampling and resampling (SISR) and martingale representations. The events in this section are assumed to belong to the σ-field generated by n random variables $Y_1, \ldots, Y_n$ on a probability space $(\Omega, \mathcal{F}, P)$. Let $\mathbf{Y}_t = (Y_1, \ldots, Y_t)$ for $1 \leq t \leq n$. For direct Monte Carlo computation of $\alpha := P\{\mathbf{Y}_n \in \Gamma\}$, i.i.d. random vectors $\mathbf{Y}_n^{(1)}, \ldots, \mathbf{Y}_n^{(m)}$ are generated from P and α is estimated by

$$(2.1) \qquad \widehat{\alpha}_{\mathrm{D}} = m^{-1} \sum_{i=1}^{m} \mathbf{1}_{\{\mathbf{Y}_n^{(i)} \in \Gamma\}}.$$

The estimate $\widehat{\alpha}_{\mathrm{D}}$ is unbiased and its variance is $\alpha(1-\alpha)/m$, which can be consistently estimated by

$$(2.2) \qquad \widehat{\sigma}_{\mathrm{D}}^2 := \widehat{\alpha}_{\mathrm{D}}(1 - \widehat{\alpha}_{\mathrm{D}})/m.$$

In most stochastic models of practical interest, the Y_t are either independent or are specified by the conditional densities $p_t(\cdot|\mathbf{Y}_{t-1})$ of Y_t given $\mathbf{Y}_{t-1}$, with respect to some measure ν. Direct Monte Carlo computation of $P\{\mathbf{Y}_n \in \Gamma\}$, therefore, involves $Y_1^{(i)}, \ldots, Y_n^{(i)}$ that are generated sequentially from these conditional densities for $1 \leq i \leq m$. In contrast, SISR first generates m independent random variables $\widetilde{Y}_t^{(1)}, \ldots, \widetilde{Y}_t^{(m)}$ at stage t, with $\widetilde{Y}_t^{(i)}$ having density function $\widetilde{q}_t(\cdot|\mathbf{Y}_{t-1}^{(i)})$, to form $\widetilde{\mathbf{Y}}_t^{(i)} = (\mathbf{Y}_{t-1}^{(i)}, \widetilde{Y}_t^{(i)})$ and then uses resampling weights of the form $w_t(\widetilde{\mathbf{Y}}_t^{(i)})/\sum_{j=1}^{m} w_t(\widetilde{\mathbf{Y}}_t^{(j)})$ to draw m independent sample paths $\mathbf{Y}_t^{(j)}$, $1 \leq j \leq m$, from $\{\widetilde{\mathbf{Y}}_t^{(i)}, 1 \leq i \leq m\}$. Here $\widetilde{q}_t$ are conditional density functions with respect to ν such that $\widetilde{q}_t > 0$ whenever $p_t > 0$; one particular choice is $\widetilde{q}_t = p_t$. In Section 3, we show how the weights w_t can be chosen to obtain logarithmically efficient SISR estimates of rare event probabilities.

The preceding SISR procedure uses *bootstrap resampling* that chooses i.i.d. sample paths from a weighted empirical measure of $\{\widetilde{\mathbf{Y}}_t^{(i)}, 1 \leq i \leq m\}$. It is, therefore, similar to the selection step of the IPS approach, that chooses

i.i.d. "path-particles" from some weighted empirical particle measure; see [14]. The Monte Carlo estimate of α using SISR with bootstrap resampling is

$$(2.3) \qquad \widehat{\alpha}_{\mathrm{B}} = m^{-1} \sum_{i=1}^{m} L(\widetilde{\mathbf{Y}}_n^{(i)}) h_{n-1}(\mathbf{Y}_{n-1}^{(i)}) \mathbf{1}_{\{\widetilde{\mathbf{Y}}_n^{(i)} \in \Gamma\}},$$

where $h_0 \equiv 1$ and

$$(2.4) \quad L(\mathbf{y}_n) = \prod_{t=1}^{n} \frac{p_t(y_t|\mathbf{y}_{t-1})}{\widetilde{q}_t(y_t|\mathbf{y}_{t-1})}, \quad h_k(\mathbf{y}_k) = \prod_{t=1}^{k} \frac{\bar{w}_t}{w_t(\mathbf{y}_t)}, \quad \bar{w}_t = \frac{1}{m} \sum_{i=1}^{m} w_t(\widetilde{\mathbf{Y}}_t^{(i)}).$$

Chan and Lai [9] have recently developed a general theory of sequential Monte Carlo filters in hidden Markov models by using a representation similar to the right-hand side of (2.3) for these filters. The method of their analysis can be applied to analyze $m(\widehat{\alpha}_{\mathrm{B}} - \alpha)$, decomposing it into a sum of $(2n-1)m$ terms so that the summands form a martingale difference sequence. Let E^* denote expectation under the probability measure $\widetilde{Q}$ from which the $\widetilde{\mathbf{Y}}_t^{(i)}$ and $\mathbf{Y}_t^{(i)}$ are drawn, and define for $1 \le t < n$,

$$(2.5) \quad f_t(\mathbf{y}_t) = E^*[L(\mathbf{Y}_n)\mathbf{1}_{\{\mathbf{Y}_n \in \Gamma\}}|\mathbf{Y}_t = \mathbf{y}_t] = L(\mathbf{y}_t)P(\mathbf{Y}_n \in \Gamma|\mathbf{Y}_t = \mathbf{y}_t),$$

setting $f_0 \equiv \alpha$ and $f_n(\widetilde{\mathbf{Y}}_n) = L(\widetilde{\mathbf{Y}}_n)\mathbf{1}_{\{\widetilde{\mathbf{Y}}_n \in \Gamma\}}$. An important ingredient in the analysis is the "ancestral origin" $a_t^{(i)}$ of $\mathbf{Y}_t^{(i)}$. Specifically, recall that the "first generation" of the m particles consists of $\widetilde{Y}_1^{(1)}, \ldots, \widetilde{Y}_1^{(m)}$ (before resampling) and set $a_t^{(i)} = j$ if the first component of $\mathbf{Y}_t^{(i)}$ is $\widetilde{Y}_1^{(j)}$. Let $\#_k^{(i)}$ denote the number of copies of $\widetilde{\mathbf{Y}}_k^{(i)}$ generated from $\{\widetilde{\mathbf{Y}}_k^{(1)}, \ldots, \widetilde{\mathbf{Y}}_k^{(m)}\}$ to form the m particles in the kth generation, and let $w_k^{(i)} = w_k(\widetilde{\mathbf{Y}}_k^{(i)})/\sum_{j=1}^{m} w_k(\widetilde{\mathbf{Y}}_k^{(j)})$. Then it follows from (2.4) and simple algebra that for $1 \le i \le m$,

$$mw_t^{(i)} = h_{t-1}(\mathbf{Y}_{t-1}^{(i)})/h_t(\widetilde{\mathbf{Y}}_t^{(i)}),$$

$$\sum_{i:a_t^{(i)}=j} f_t(\mathbf{Y}_t^{(i)}) h_t(\mathbf{Y}_t^{(i)}) = \sum_{i:a_{t-1}^{(i)}=j} \#_t^{(i)} f_t(\widetilde{\mathbf{Y}}_t^{(i)}) h_t(\widetilde{\mathbf{Y}}_t^{(i)}),$$

$$\sum_{t=1}^{n} \sum_{i:a_{t-1}^{(i)}=j} [f_t(\widetilde{\mathbf{Y}}_t^{(i)}) - f_{t-1}(\mathbf{Y}_{t-1}^{(i)})] h_{t-1}(\mathbf{Y}_{t-1}^{(i)})$$

$$+ \sum_{t=2}^{n} \sum_{i:a_{t-2}^{(i)}=j} (\#_{t-1}^{(i)} - mw_{t-1}^{(i)}) f_{t-1}(\widetilde{\mathbf{Y}}_{t-1}^{(i)}) h_{t-1}(\widetilde{\mathbf{Y}}_{t-1}^{(i)})$$

$$= \sum_{i:a_{n-1}^{(i)}=j} f_n(\widetilde{\mathbf{Y}}_n^{(i)})h_{n-1}(\mathbf{Y}_{n-1}^{(i)}) - \alpha,$$

recalling that $f_0 \equiv \alpha$, $h_0 \equiv 1$ and defining $a_0^{(i)} = i$. Let

$$(2.6)\quad \epsilon_{2t-1}^{(j)} = \sum_{i:a_{t-1}^{(i)}=j} [f_t(\widetilde{\mathbf{Y}}_t^{(i)}) - f_{t-1}(\mathbf{Y}_{t-1}^{(i)})]h_{t-1}(\mathbf{Y}_{t-1}^{(i)}) \text{ for } 1 \le t \le n,$$

$$\epsilon_{2t}^{(j)} = \sum_{i:a_{t-1}^{(i)}=j} (\#_t^{(i)} - mw_t^{(i)})[f_t(\widetilde{\mathbf{Y}}_t^{(i)})h_t(\widetilde{\mathbf{Y}}_t^{(i)}) - \alpha] \text{ for } 1 \le t \le n-1.$$

Then for each fixed j, $\{\epsilon_t^{(j)}, 1 \le t \le 2n-1\}$ is a martingale difference sequence with respect to the filtration $\{\mathcal{F}_t, 1 \le t \le 2n-1\}$ defined below, and

$$(2.7)\qquad\qquad m(\widehat{\alpha}_{\mathrm{B}} - \alpha) = \sum_{j=1}^{m}(\epsilon_1^{(j)} + \cdots + \epsilon_{2n-1}^{(j)}).$$

The martingale representation (2.7) that involves the ancestral origins of the genealogical particles is useful for estimating the standard error of $\widehat{\alpha}_{\mathrm{B}}$, as shown by Chan and Lai [9] who have also introduced the σ-fields

$$(2.8)\quad \begin{aligned} \mathcal{F}_{2t-1} &= \sigma(\{\widetilde{Y}_1^{(i)} : 1 \le i \le m\} \\ &\qquad \cup \{(\mathbf{Y}_s^{(i)}, \widetilde{\mathbf{Y}}_{s+1}^{(i)}, a_s^{(i)}) : 1 \le s < t, 1 \le i \le m\}), \\ \mathcal{F}_{2t} &= \sigma(\mathcal{F}_{2t-1} \cup \{(\mathbf{Y}_t^{(i)}, a_t^{(i)}) : 1 \le i \le m\}), \end{aligned}$$

with respect to which (2.6) forms a martingale difference sequence.

Since $f_n(\widetilde{\mathbf{Y}}_n^{(i)}) = L(\widetilde{\mathbf{Y}}_n^{(i)})\mathbf{1}_{\{\widetilde{\mathbf{Y}}_n^{(i)} \in \Gamma\}}$ and $\sum_{i=1}^{m}(\#_t^{(i)} - mw_t^{(i)}) = 0$ for $1 \le t \le n-1$, summing (2.6) over t and j yields (2.7). Without tracing their ancestral origins, we can also use the successive generations of the m particles to form martingale differences directly. Specifically, in analogy with (2.6), define for $i = 1, \dots, m$,

$$(2.9)\quad Z_{2t-1}^{(i)} = [f_t(\widetilde{\mathbf{Y}}_t^{(i)}) - f_{t-1}(\mathbf{Y}_{t-1}^{(i)})]h_{t-1}(\mathbf{Y}_{t-1}^{(i)}) \text{ for } 1 \le t \le n,$$

$$Z_{2t}^{(i)} = f_t(\mathbf{Y}_t^{(i)})h_t(\mathbf{Y}_t^{(i)}) - \sum_{j=1}^{m} w_t^{(j)} f_t(\widetilde{\mathbf{Y}}_t^{(j)})h_t(\widetilde{\mathbf{Y}}_t^{(j)}) \text{ for } 1 \le t \le n-1.$$

As noted by Chan and Lai [9], $\{(Z_t^{(1)}, \dots, Z_t^{(m)}), 1 \le t \le 2n-1\}$ is a martingale difference sequence with respect to the filtration $\{\mathcal{F}_t, 1 \le t \le 2n-1\}$

and $Z_t^{(1)}, \ldots, Z_t^{(m)}$ are conditionally independent given $\mathcal{F}_{t-1}$; moreover,

$$(2.10) \qquad m(\widehat{\alpha}_{\mathrm{B}} - \alpha) = \sum_{t=1}^{2n-1} (Z_t^{(1)} + \cdots + Z_t^{(m)}).$$

From the martingale representation (2.10), it follows that $E^*(\widehat{\alpha}_{\mathrm{B}}) = \alpha$. Moreover, under the assumption that

$$(2.11) \quad \sigma_{\mathrm{B}}^2 := \sum_{t=1}^{n} E^* \Big[f_t^2(\mathbf{Y}_t) \Big/ \prod_{k=1}^{t-1} w_k(\mathbf{Y}_k) \Big] E^* \Big[\prod_{k=1}^{t-1} w_k(\mathbf{Y}_k) \Big] - n\alpha^2 < \infty,$$

application of the central limit theorem yields

$$(2.12) \qquad \sqrt{m}(\widehat{\alpha}_B - \alpha) \Rightarrow N(0, \sigma_{\mathrm{B}}^2) \text{ as } m \to \infty.$$

A consistent estimate of σ_{B}^2 is given by

$$(2.13) \qquad \widehat{\sigma}_{\mathrm{B}}^2 \;\; := \;\; m^{-1} \sum_{j=1}^{m} \Big\{ \Big[\sum_{i:a_{n-1}^{(i)}=j} f_n(\widetilde{\mathbf{Y}}_n^{(i)}) h_{n-1}(\mathbf{Y}_{n-1}^{(i)}) \Big]$$
$$- \Big[1 + \sum_{t=1}^{n-1} \sum_{i:a_{t-1}^{(i)}=j} (\#_t^{(i)} - m w_t^{(i)}) \Big] \widehat{\alpha}_{\mathrm{B}} \Big\}^2,$$

which can be shown to converge to σ_{B}^2 in probability as $m \to \infty$ by making use of the martingale representation (2.7); see [9] for details. Del Moral and Jacod [15] have derived by a different method a martingale representation similar to (2.10); see (3.3.7) and (3.3.8) of [15], in which the term $Z_{2t-1}^{(i)}$ in (2.9) corresponds to the tth mutation on the ith particle and $Z_{2t}^{(i)}$ the tth selection by the ith particle. In [15], these two terms are combined into a sum, and a central limit theorem similar to (2.12) is proved under the assumption of bounded f_n.

Note that in (2.12) on the asymptotic normality of $\widehat{\alpha}_{\mathrm{B}}$ and in the consistency result $\widehat{\sigma}_{\mathrm{B}}^2 \xrightarrow{p} \sigma_{\mathrm{B}}^2$, the sample size n in the probability $\alpha = P\{\mathbf{Y}_n \in \Gamma\}$ is assumed to be fixed whereas the number m of Monte Carlo samples approaches ∞. The consistent estimate $\widehat{\sigma}_{\mathrm{B}}^2$ of σ_{B}^2 in (2.13) provides an estimate $\widehat{\sigma}_{\mathrm{B}}/\sqrt{m}$ of the standard error s.e.$(\widehat{\alpha}_{\mathrm{B}})$ of the Monte Carlo estimate $\widehat{\alpha}_{\mathrm{B}}$. Note that the usual estimate $\sqrt{\widehat{\alpha}_{\mathrm{B}}(1 - \widehat{\alpha}_{\mathrm{B}})}$ is inconsistent for $\sqrt{m}$ s.e.$(\widehat{\alpha}_{\mathrm{B}})$ because of the dependence among the m sample paths due to resampling in the SISR procedure as in [13, 14]. The case of n approaching ∞ will be considered in the next section, in which the representation (2.6) will still play a pivotal

role, but which requires new methods and large deviation principles rather than central limit theorems

Instead of bootstrap resampling, we can use the residual resampling scheme introduced by Baker [1, 2] which often leads to smaller asymptotic variance than that of bootstrap resampling. We consider here a variant of this scheme, introduced by Crisan, Del Moral and Lyons [11], that can result in further reduction of the asymptotic variance. Let $\lfloor \cdot \rfloor$ denote the greatest integer function and let m_t be the sample size at stage t, with $m_1 = m$. We modify the bootstrap resampling step of the SISR procedure as follows: Let $U_t^{(1)}, \ldots, U_t^{(m_t)}$ be independent Bernoulli random variables satisfying $P\{U_t^{(i)} = 1\} = m_t w_t^{(i)} - \lfloor m_t w_t^{(i)} \rfloor$. For each $1 \leq i \leq m_t$ and $t < n$, make $\#_t^{(i)} := \lfloor m_t w_t^{(i)} \rfloor + U_t^{(i)}$ copies of $(\widetilde{\mathbf{Y}}_t^{(i)}, a_{t-1}^{(i)}, h_{t-1}^{(i)}, w_t^{(i)})$. These copies constitute an augmented sample $\{(\mathbf{Y}_t^{(j)}, a_t^{(j)}, h_t^{(j)}, w_t^{(j)}) : 1 \leq j \leq m_{t+1}\}$, where $m_{t+1} = \sum_{i=1}^{m_t} \#_t^{(i)}$ and $h_t^{(i)} = h_{t-1}^{(i)}/(m_t w_t^{(i)})$. Estimate α by

$$\widehat{\alpha}_{\mathrm{R}} := m_n^{-1} \sum_{i=1}^{m_n} L(\widetilde{\mathbf{Y}}_n^{(i)}) h_{n-1}^{(i)}(\mathbf{Y}_{n-1}^{(i)}) \mathbf{1}_{\{\widetilde{\mathbf{Y}}_n^{(i)} \in \Gamma\}}.$$

Define $\epsilon_k^{(j)}$ by (2.6) in which m is replaced by m_t, and define $\mathcal{F}_{2t-1}$ (or $\mathcal{F}_{2t}$) by (2.8), in which m is replaced by m_{s+1} (or by m_{t+1}). Moreover, define

$$\begin{aligned}
\widetilde{Z}_{2t-1}^{(i)} &= [f_t(\widetilde{\mathbf{Y}}_t^{(i)}) - f_{t-1}(\mathbf{Y}_{t-1}^{(i)})]h_{t-1}(\mathbf{Y}_{t-1}^{(i)}) \text{ for } 1 \leq t \leq n, \\
\widetilde{Z}_{2t}^{(i)} &= (\#_t^{(i)} - m_t w_t^{(i)})[f_t(\widetilde{\mathbf{Y}}_t^{(i)})h_t(\widetilde{\mathbf{Y}}_t^{(i)}) - \alpha] \text{ for } 1 \leq t \leq n-1,
\end{aligned}$$

for $i = 1, \ldots, m_t$. Recall that the first generation of particles consists of $\widetilde{Y}_1^{(1)}, \ldots, \widetilde{Y}_1^{(m)}$ and that $a_t^{(i)} = j$ if the first component of $\mathbf{Y}_t^{(i)}$ is $\widetilde{Y}_1^{(j)}$ for $j = 1, \ldots, m$ and $i = 1, \ldots, m_{t+1}$. Analogous to (2.7) and (2.10), we have the martingale representations

$$(2.14) \quad m_n(\widehat{\alpha}_{\mathrm{R}} - \alpha) = \sum_{j=1}^{m}(\epsilon_1^{(j)} + \cdots + \epsilon_{2n-1}^{(j)}) = \sum_{k=1}^{2n-1}(\widetilde{Z}_k^{(1)} + \cdots + \widetilde{Z}_k^{(m_{\lfloor (k+1)/2 \rfloor})}).$$

Analogous to (2.13), define

$$\begin{aligned}
\widehat{\sigma}_{\mathrm{R}}^2 &= m^{-1} \sum_{j=1}^{m} \left\{ \left[\sum_{i:a_{n-1}^{(i)}=j} f_n(\widetilde{\mathbf{Y}}_n^{(i)}) h_{n-1}(\mathbf{Y}_{n-1}^{(i)}) \right] \right. \\
&\quad \left. - \left[1 + \sum_{t=1}^{n-1} \sum_{i:a_{t-1}^{(i)}=j} (\#_t^{(i)} - m_t w_t^{(i)}) \right] \widehat{\alpha}_{\mathrm{R}} \right\}^2.
\end{aligned}$$

From (2.14), it follows that $E^*[m_n(\widehat{\alpha}_{\mathrm{R}} - \alpha)] = 0$. Let

$$\eta_t = E^*\Big[\prod_{k=1}^{t} w_k(\mathbf{Y}_k)\Big], \qquad h_t^*(\mathbf{y}_t) = \eta_t \Big/ \prod_{k=1}^{t} w_k(\mathbf{y}_k),$$

and let $\gamma(x) = (x - \lfloor x \rfloor)(1 - x + \lfloor x \rfloor)/x$ for $x > 0$. If (2.11) holds, then analogous to corresponding results for $\widehat{\alpha}_{\mathrm{B}}$ and $\widehat{\sigma}_{\mathrm{B}}^2$ in the bootstrap resampling case, we now have as $m \to \infty$,

$$\widehat{\sigma}_{\mathrm{R}}^2 \xrightarrow{p} \sigma_{\mathrm{R}}^2, \quad m_t/m \xrightarrow{p} 1 \text{ for every } t \geq 1,$$

$$\sqrt{m}(\widehat{\alpha}_{\mathrm{R}} - \alpha) \Rightarrow N(0, \sigma_{\mathrm{R}}^2),$$

where $\sigma_{\mathrm{R}}^2 < \sigma_{\mathrm{B}}^2$ and

$$\sigma_{\mathrm{R}}^2 \;\; := \;\; \sum_{t=1}^{n} E^*\{[f_t^2(\mathbf{Y}_t) - f_{t-1}^2(\mathbf{Y}_{t-1})]h_{t-1}^*(\mathbf{Y}_{t-1})\}$$
$$+ \sum_{t=1}^{n-1} E^*\Big\{\gamma\Big(\frac{h_{t-1}^*(\mathbf{Y}_{t-1})}{h_t^*(\mathbf{Y}_t)}\Big)\frac{[f_t(\mathbf{Y}_t)h_t^*(\mathbf{Y}_t) - \alpha]^2}{h_t^*(\mathbf{Y}_t)}\Big\}.$$

Details are given in [9]. Note the additional variance reduction if residual resampling is used instead of bootstrap resampling.

3. Logarithmically efficient SISR for Monte Carlo computation of small tail probabilities. Let $\xi, \xi_1, \xi_2, \ldots$ be i.i.d. d-dimensional random vectors with a common distribution function F such that $\psi(\theta) := \log(Ee^{\theta'\xi}) < \infty$ for $\|\theta\| < \theta_0$. Let $S_n = \xi_1 + \cdots + \xi_n$, $\mu_0 = E\xi$, $\Theta = \{\theta : \psi(\theta) < \infty\}$, and let Λ be the closure of $\nabla\psi(\Theta)$ and Λ^o be its interior. Assume that for any $\theta_0 \in \Theta^o$ and $\theta \in \Theta \setminus \Theta^o$,

$$\lim_{\rho\uparrow 1}(\theta - \theta_0)'\nabla\psi(\theta_0 + \rho(\theta - \theta_0)) = \infty.$$

Then by convex analysis, (see e.g. [4, Chapter 3]), Λ contains the convex hull of the support of $\{S_n/n, n \geq 1\}$. The gradient vector $\nabla\psi$ is a diffeomorphism from Θ^o onto Λ^o. For given $\mu \in \Lambda^o$, let $\theta_\mu = (\nabla\psi)^{-1}(\mu)$ and define the *rate function*

(3.1) $$\phi(\mu) = \sup_{\theta \in \Theta}\{\theta'\mu - \psi(\theta)\} = \theta_\mu'\mu - \psi(\theta_\mu).$$

We can embed F in an exponential family $\{F_\theta, \theta \in \Theta\}$ with $dF_\theta(x) = e^{\theta'x - \psi(\theta)}dF(x)$. Under certain regularity conditions on $g : \Lambda \to \mathbf{R}$, Chan

and Lai [6] have developed asymptotic approximations, which involve both g and ϕ, to the exceedance probabilities

$$(3.2) \qquad p_n \quad = \quad P\{g(S_n/n) \geq b\} \text{ with } b > g(\mu_0),$$

$$(3.3) \qquad p_c \quad = \quad P\{\max_{n_0 \leq n \leq n_1} ng(S_n/n) \geq c\},$$

where $n_0 \sim \rho_0 c$ and $n_1 \sim \rho_1 c$ such that $g(\mu_0) < \rho_1^{-1}$. Making use of these approximations, Chan and Lai [8] have shown that certain mixtures of exponentially twisted measures are asymptotically optimal for Monte Carlo evaluation of (3.2) or (3.3) by importance sampling. Specifically, for $A = \{g(S_n/n) \geq b\}$ in the case of (3.2) or $A = \{\max_{n_0 \leq n \leq n_1} ng(S_n/n) \geq c\}$ in the case of (3.3), an importance sampling measure Q (which may depend on n or c) is said to be *asymptotically optimal* if

$$(3.4) \qquad m\text{Var}\Big(m^{-1}\sum_{i=1}^{m} L_i \mathbf{1}_{A_i}\Big) = O(\sqrt{n}p_n^2) \text{ as } n \to \infty$$

in the case of (3.2), and if

$$(3.5) \qquad m\text{Var}\Big(m^{-1}\sum_{i=1}^{m} L_i \mathbf{1}_{A_i}\Big) = O(p_c^2) \text{ as } c \to \infty$$

in the case of (3.3), where $(L_1, \mathbf{1}_{A_1}), \ldots, (L_m, \mathbf{1}_{A_m})$ are m independent realizations of $(L := dP/dQ, \mathbf{1}_A)$. For the case of (3.3), since $E_Q(L\mathbf{1}_A) = P(A) = p_c$, $E_Q(L^2\mathbf{1}_A) \geq p_c^2$ by the Cauchy-Schwarz inequality and therefore Q is an asymptotically optimal importance sampling measure if $E_Q(L^2\mathbf{1}_A) = O(p_c^2)$, which leads to the definition (3.5) of asymptotic optimality for the Monte Carlo estimates. Chan and Lai [8] have also shown that $\sqrt{n}p_n^2$ is an asymptotically minimal order of magnitude for $E_Q(L^2\mathbf{1}_A)$ in the case of (3.2). They have also extended this theory to Markov random walks S_n whose increments ξ_i have distributions $F(\cdot|X_i, X_{i-1})$ depending on a Markov chain $\{X_t\}$.

The asymptotically optimal mixtures of exponentially twisted measures $\int P_{\theta_\mu}\omega(\mu)d\mu$ in [8] involve normalizing constants β_n (or β_c) that may be difficult to compute. Moreover, it may even be difficult to sample from the twisted measure P_{θ_μ}, especially in multidimensional and Markovian settings. In this section we show that by choosing the resampling weights suitably, the SISR estimates $\widehat{\alpha}_{\mathrm{B}}$ can still attain

$$(3.6) \qquad m\text{Var}(\widehat{\alpha}_{\mathrm{B}}) = p_n^2 e^{o(n)} \text{ as } m \to \infty \text{ and } n \to \infty$$

for Monte Carlo estimation of p_n, and

$$(3.7) \qquad m\mathrm{Var}(\widehat{\alpha}_{\mathrm{B}}) = p_c^2 e^{o(c)} \text{ as } m \to \infty \text{ and } c \to \infty$$

for Monte Carlo estimation of (3.3). Moreover, (3.6) and (3.7) still hold with $\widehat{\alpha}_{\mathrm{B}}$ replaced by $\widehat{\alpha}_{\mathrm{R}}$. The properties (3.6) and (3.7) are called *logarithmic efficiency*; the variance of the Monte Carlo estimate differs from the asymptotically optimal value by a factor of $e^{o(n)}$ (or $e^{o(c)}$), noting that $-n^{-1}\log p_n$ and $-c^{-1}\log p_c$ converge to positive limits. To begin with, suppose the asymptotically optimal importance sampling measure Q has conditional densities $q_t(\cdot|\mathbf{Y}_{t-1})$ with respect to ν. To achieve log efficiency, the resampling functions w_t can be chosen to satisfy approximately

$$(3.8) \qquad w_t(\mathbf{y}_t) \propto q_t(y_t|\mathbf{y}_{t-1})/\widetilde{q}_t(y_t|\mathbf{y}_{t-1}),$$

as illustrated by the following, after which a heuristic explanation for (3.8) will be given.

EXAMPLE 1. Suppose $\xi_1, \xi_2, \ldots$ are i.i.d. random variables ($d = 1$) and $g(x) = x$ in (3.2), so that $\alpha = p_n = P\{S_n/n \geq b\}$, where $b > E\xi_1$ and $2\theta_b \in \Theta$. Consider the SISR procedure with $\widetilde{Q} = P$ (and therefore $E^* = E$) and resampling weights

$$(3.9) \qquad w_t(\mathbf{Y}_t) = e^{\theta_b \xi_t - \psi(\theta_b)}.$$

Then $L = 1$ and hence by (2.5),

$$(3.10) \qquad f_t(\mathbf{Y}_t) = P\{S_n/n \geq b|\mathbf{Y}_t\} = P\{S_n - S_t \geq nb - S_t|S_t\}.$$

Therefore standard Markov's inequality involving moment generating functions yields

$$(3.11) \qquad f_t(\mathbf{Y}_t) \leq e^{-\theta_b(nb-S_t)+(n-t)\psi(\theta_b)} = e^{\theta_b S_t - t\psi(\theta_b) - n\phi(b)}.$$

By (2.6) and the martingale decomposition (2.7),

$$(3.12) \quad E(\widehat{\alpha}_{\mathrm{B}} - \alpha)^2 \leq m^{-1} \sum_{t=1}^{n} E\{[f_t(\widetilde{\mathbf{Y}}_t^{(1)}) - f_{t-1}(\mathbf{Y}_{t-1}^{(1)})]^2 h_{t-1}^2(\mathbf{Y}_{t-1}^{(1)})\}$$

$$+ m^{-1} \sum_{t=1}^{n-1} E[(\#_t^{(1)} - mw_t^{(1)})^2 f_t^2(\widetilde{\mathbf{Y}}_t^{(1)}) h_t^2(\widetilde{\mathbf{Y}}_t^{(1)})],$$

in which the superscript $^{(1)}$ can be replaced by $^{(i)}$ since the expectations are the same for all i. The derivation of (3.12) uses the independence of

$[f_t(\widetilde{\mathbf{Y}}_t^{(i)}) - f_{t-1}(\mathbf{Y}_{t-1}^{(i)})]h_t(\mathbf{Y}_{t-1}^{(i)})$ for $1 \le i \le m$ when conditioned on $\mathcal{F}_{2t-2}$ and the pairwise negative correlations of $(\#_t^{(i)} - mw_t^{(i)})f_t(\widetilde{\mathbf{Y}}_t^{(i)})h_t(\widetilde{\mathbf{Y}}_t^{(i)})$ for $i = 1, \ldots, m$ when conditioned on $\mathcal{F}_{2t-1}$. By (2.4), (3.9) and (3.11),

$$(3.13) \quad E\{[f_t(\widetilde{\mathbf{Y}}_t^{(1)}) - f_{t-1}(\mathbf{Y}_{t-1}^{(1)})]^2 h_{t-1}^2(\mathbf{Y}_{t-1}^{(1)})\}$$

$$= E\{\bar{w}_1^2 \cdots \bar{w}_{t-1}^2 [f_t(\widetilde{\mathbf{Y}}_t^{(1)}) - f_{t-1}(\mathbf{Y}_{t-1}^{(1)})]^2 / e^{2\theta_b S_{t-1}^{(1)} - 2(t-1)\psi(\theta_b)}\}$$

$$\le \Big(1 + \frac{E(e^{\theta_b \xi_1 - \psi(\theta_b)} - 1)^2}{m}\Big)^{t-1} e^{-2n\phi(b)} E(e^{2\theta_b \xi_t - 2\psi(\theta_b)}).$$

To see the inequality in (3.13), condition on $\mathcal{F}_{2t-1}$. Since $E[f_t(\widetilde{\mathbf{Y}}_t^{(1)})|\mathcal{F}_{2t-1}] = f_{t-1}(\mathbf{Y}_{t-1}^{(1)})$, it follows from (3.11) that

$$E\{[f_t(\widetilde{\mathbf{Y}}_t^{(1)}) - f_{t-1}(\mathbf{Y}_{t-1}^{(1)})]^2 / e^{2\theta_b S_{t-1}^{(1)} - 2(t-1)\psi(\theta_b)}|\mathcal{F}_{2t-1}]$$

$$\le E[f_t^2(\widetilde{\mathbf{Y}}_t^{(1)}) / e^{2\theta_b S_{t-1}^{(1)} - 2(t-1)\psi(\theta_b)}|\mathcal{F}_{2t-1}] \le e^{-2n\phi(b)} E(e^{2\theta_b \xi_t - 2\psi(\theta_b)}).$$

Moreover, $\bar{w}_1^2, \ldots, \bar{w}_{t-1}^2$ are i.i.d. random variables with mean

$$(3.14) \quad E\Big[m^{-1} \sum_{i=1}^{m} (e^{\theta_b \xi_1^{(i)} - \psi(\theta_b)} - 1) + 1\Big]^2 = 1 + m^{-1} E(e^{\theta_b \xi_1 - \psi(\theta_b)} - 1)^2,$$

and their product $\bar{w}_1^2 \cdots \bar{w}_{t-1}^2$ in the second term of (3.13) is $\mathcal{F}_{2t-1}$-measurable. This yields the inequality in (3.13).

Since the conditional distribution of $\#_t^{(i)}$ given $\mathcal{F}_{2t-1}$ is Binomial$(m, w_t^{(i)})$, $E[(\#_t^{(i)} - mw_t^{(i)})^2|\mathcal{F}_{2t-1}] \le mw_t^{(i)}$. By (2.4), (3.9) and (3.11), $f_t(\widetilde{\mathbf{Y}}_t^{(i)})h_t(\widetilde{\mathbf{Y}}_t^{(i)}) \le \bar{w}_1 \cdots \bar{w}_t e^{-n\phi(b)}$. Since $\sum_{i=1}^{m} w_t^{(i)} = 1$, it then follows by conditioning on $\mathcal{F}_{2t-1}$ that

$$E\{(\#_t^{(1)} - mw_t^{(1)})^2 f_t^2(\widetilde{\mathbf{Y}}_t^{(1)})h_t^2(\widetilde{\mathbf{Y}}_t^{(1)})\}$$

$$= m^{-1} \sum_{i=1}^{m} E\{(\#_t^{(i)} - mw_t^{(i)})^2 f_t^2(\widetilde{\mathbf{Y}}_t^{(i)})h_t^2(\widetilde{\mathbf{Y}}_t^{(i)})\}$$

$$\le E\Big\{\Big(\sum_{i=1}^{m} w_t^{(i)}\Big)(\bar{w}_1 \cdots \bar{w}_t e^{-n\phi(b)})^2\Big\} = e^{-2n\phi(b)} E(\bar{w}_1^2 \cdots \bar{w}_t^2),$$

which can be combined with (3.14) to yield

$$(3.15) \quad E[(\#_t^{(1)} - mw_t^{(1)})^2 f_t^2(\widetilde{\mathbf{Y}}_t^{(1)})h_t^2(\widetilde{\mathbf{Y}}_t^{(1)})] = O\Big(\Big(1 + \frac{K}{m}\Big)^t e^{-2n\phi(b)}\Big),$$

where $K = E(e^{\theta_b \xi_1 - \psi(\theta_b)} - 1)^2$. By (3.12), (3.13) and (3.15),

$$\liminf_{n \to \infty} -\frac{1}{n} \log[m \mathrm{Var}(\widehat{\alpha}_\mathrm{B})] \ge 2\phi(b) - \frac{K}{m}$$

for any fixed m. Since $p_n/[n^{-1/2} e^{-n\phi(b)}]$ is bounded away from 0 and ∞ (see [8, p.451]), (3.6) holds.

3.1. *A heuristic principle for efficient SISR procedures.* The asymptotically optimal importance sampling measure for $p_n = P\{S_n/n \geq b\}$ is Q under which $\xi_1, \xi_2, \ldots$ are i.i.d. with density function $e^{\theta_b \xi - \psi(\theta_b)}$ with respect to P; see [8]. Since we have used $\widetilde{Q} = P$ in Example 1, (3.9) actually follows the prescription (3.8) to choose resampling weights that can achieve an effect similar to asymptotically optimal importance sampling. We now give a heuristic principle underlying this prescription. The SISR procedure uses the importance weights $p_t^{(i)}/\widetilde{q}_t^{(i)}$ (for the change of measures from P to $\widetilde{Q}$) and resampling weights $w_t^{(i)}$, $1 \leq i \leq m$, for the m simulated trajectories at stage t. The resampling step at stage t basically converts $(\widetilde{\mathbf{Y}}_t^{(i)}, p_t^{(i)}/\widetilde{q}_t^{(i)}, w_t^{(i)})$ to $(\mathbf{Y}_t^{(i)}, p_t^{(i)}/(\widetilde{q}_t^{(i)} w_t^{(i)}), 1)$, and therefore the prescription (3.8) for choosing resampling weights (satisfying $\widetilde{q}_t^{(i)} w_t^{(i)} = q_t^{(i)}$) is intended to yield the desired importance weights $p_t^{(i)}/q_t^{(i)}$. To transform this heuristic principle into a rigorous proof of logarithmic efficiency, one needs to be able to bound the second moments of the importance weights and resampling weights. This explains the requirement $2\theta_b \in \Theta$ in Example 1.

Example 1 indicates the key role played by the martingale decomposition (2.7) and large deviation bounds for $P(\Gamma_n|\mathbf{Y}_k)$, $1 \leq k < n$, in the derivation of asymptotically efficient resampling weights. To generalize the basic ideas to the more general tail probability (3.2) with nonlinear g, we provide large deviation bounds in Lemma 1, whose proof is given in the Appendix, for

$$(3.16) \qquad P\{g((x + S_{n,k})/n) \geq b\},$$

where $S_{n,k} = S_n - S_k$; note that (3.16) is equal to $P\{g(S_n/n) \geq b|S_k = x\}$. The special case $k = 0$ and $x = 0$ has been analyzed by Chan and Lai (see Theorem 2 of [6]) under certain regularity conditions that yield precise saddlepoint approximations. The probability (3.16) is more complicated than this special case because it involves additional parameters x and k, but we only need large deviation bounds rather than saddlepoint approximations for logarithmic efficiency. Let $\mu_\theta = \nabla \psi(\theta)$ and define

$$(3.17) \qquad I \quad = \quad \inf\{\phi(\mu) : g(\mu) \geq b\},$$
$$(3.18) \qquad M \quad = \quad \{\theta : \phi(\mu_\theta) \leq I\}.$$

LEMMA 1.   *Let $b > g(\mu_0)$. Then as $n \to \infty$,*

$$(3.19) \qquad P\{g((x + S_{n,k})/n) \geq b\} \leq e^{-nI + o(n)} \int_M e^{\theta' x - k\psi(\theta)} d\theta,$$

where the $o(n)$ term is uniform in x and k.

The proof of (3.19) in the Appendix uses a change-of-measure argument that involves the measure Q for which

$$(dQ/dP)(\mathbf{Y}_n) = \int_M e^{\theta' S_n - n\psi(\theta)} d\theta \Big/ \mathrm{vol}(M).$$

The bound (3.19) is used in conjunction with the inequality $\int_M e^{\theta' x - k\psi(\theta)} d\theta \leq \mathrm{vol}(M) \exp\{k \max_{\theta \in M}[\theta' x/k - \psi(\theta)]\}$ to prove the following.

THEOREM 1. *Letting $b > g(\mu_0)$, assume*

(C1) *g is twice continuously differentiable, and $\nabla g \neq 0$ on*
       $N := \{\mu \in \Lambda^o : g(\mu) = b\}$,

(C2) *$E e^{2\kappa \|\xi_1\|} < \infty$,*

where $\kappa = \sup_{\theta \in M} \|\theta\|$ and M is defined in (3.18). Let $\widehat{\theta}_0 = 0$ and define for $1 \leq t \leq n$,

$$(3.20) \qquad \widehat{\theta}_t = \arg \max_{\theta \in M} \{\theta' S_t/t - \psi(\theta)\},$$

$$w_t(\mathbf{Y}_t) = \exp\{\widehat{\theta}_t' S_t - t\psi(\widehat{\theta}_t) - [\widehat{\theta}_{t-1}' S_{t-1} - (t-1)\psi(\widehat{\theta}_{t-1})]\}.$$

With $\widetilde{Q} = P$ and the resampling weights thus defined, the SISR estimates $\widehat{\alpha}_{\mathrm{B}}$ and $\widehat{\alpha}_{\mathrm{R}}$ are logarithmically efficient, i.e., (3.6) holds for $\widehat{\alpha}_{\mathrm{B}}$ and also with $\widehat{\alpha}_{\mathrm{R}}$ in place of $\widehat{\alpha}_{\mathrm{B}}$ if $m \to \infty$ and $n \to \infty$.

Besides (3.19), the proof of Theorem 1 also uses the bounds in the following lemma. These bounds enable us to bound $E(\bar{w}_{t-1}^2 | \mathcal{F}_{2(t-1)-2})$ in the proof of Theorem 1.

LEMMA 2. *With the same notation and assumptions in Theorem 1, there exist nonrandom constants ϵ_t and $K > 0$ such that*
$$(3.21)$$
$$\lim_{t \to \infty} \epsilon_t = 0, \quad E[w_t(\mathbf{Y}_t)|S_{t-1}] \leq e^{\epsilon_t} \text{ and } E[w_t^2(\mathbf{Y}_t)|S_{t-1}] \leq K \text{ for all } t \geq 1.$$

PROOF. Let $\eta = \sup_{\theta \in M} |\psi(\theta)|$. Then

$$(3.22) \quad \widehat{\theta}_t' S_t - t\psi(\widehat{\theta}_t) = [\widehat{\theta}_t' S_{t-1} - (t-1)\psi(\widehat{\theta}_t)] + [\widehat{\theta}_t' \xi_t - \psi(\widehat{\theta}_t)]$$
$$\leq [\widehat{\theta}_{t-1}' S_{t-1} - (t-1)\psi(\widehat{\theta}_{t-1})] + [\widehat{\theta}_t' \xi_t - \psi(\widehat{\theta}_t)],$$

and therefore it follows from (3.20) that $w_t(\mathbf{Y}_t) \leq e^{\kappa \|\xi_t\| + \eta}$. Hence, by (C2),

$$(3.23) \quad E[w_t(\mathbf{Y}_t) \mathbf{1}_{\{\|\xi_t\| > \zeta\}}|S_{t-1}] \leq E[e^{\kappa \|\xi_1\| + \eta} \mathbf{1}_{\{\|\xi_1\| > \zeta\}}] \to 0 \text{ as } \zeta \to \infty.$$

It will be shown that for any fixed $\zeta > 0$,

$$(3.24) \qquad \gamma_{t,\zeta} := \operatorname{ess\,sup} \|\widehat{\theta}_t - \widehat{\theta}_{t-1}\| \mathbf{1}_{\{\|\xi_t\| \leq \zeta\}} \to 0 \text{ as } t \to \infty.$$

Let $\widetilde{\eta} = \sup_{\theta \in M} \|\nabla \psi(\theta)\|$. Combining (3.24) with (3.20) and (3.22) yields

$$(3.25) \qquad E[w_t(\mathbf{Y}_t)\mathbf{1}_{\{\|\xi_t\| \leq \zeta\}}|S_{t-1}] \leq E[e^{\widehat{\theta}_t' \xi_t - \psi(\widehat{\theta}_t)}\mathbf{1}_{\{\|\xi_t\| \leq \zeta\}}|S_{t-1}]$$
$$\leq e^{\gamma_{t,\zeta}(\zeta+\widetilde{\eta})} E[e^{\widehat{\theta}_{t-1}' \xi_t - \psi(\widehat{\theta}_{t-1})}|S_{t-1}] = 1 + o(1)$$

as $t \to \infty$. Moreover, by (C2) and (3.24), as $\zeta \to \infty$,

$$(3.26) \quad E[w_t^2(\mathbf{Y}_t)\mathbf{1}_{\{\|\xi_t\| > \zeta\}}|S_{t-1}] \quad \leq \quad E[e^{2\kappa\|\xi_1\| + 2\eta}\mathbf{1}_{\{\|\xi_1\| > \zeta\}}] \to 0$$
$$E[w_t^2(\mathbf{Y}_t)\mathbf{1}_{\{\|\xi_t\| \leq \zeta\}}|S_{t-1}] \quad \leq \quad e^{2\gamma_{t,\zeta}(\zeta+\widetilde{\eta})} E[e^{2\widehat{\theta}_{t-1}' \xi_t - 2\psi(\widehat{\theta}_{t-1})}|S_{t-1}]$$
$$\leq \quad \sup_{\theta \in M} e^{\psi(2\theta) - 2\psi(\theta)} + o(1),$$

and (3.21) follows from (3.23), (3.25) and (3.26).

To prove (3.24), let $f_{x,t}(\theta) = \theta'x - t\psi(\theta)$ and let $\theta_{x,t}$ be the unique maximizer of $f_{x,t}(\theta)$ over M. Let $\lambda_{\min}(\cdot)$ denote the smallest eigenvalue of a symmetric matrix. Since $\nabla^2 \psi(\theta)$ is continuous and positive definite for all $\theta \in M$, and since M is compact and $\lambda_{\min}$ is a continuous function of the entries of $\nabla^2 \psi(\theta)$, $\inf_{\theta \in M} \lambda_{\min}(\nabla^2\psi(\theta)) \geq 2\beta$ for some $\beta > 0$. Therefore by Taylor's theorem, $f_{x,t-1}(\theta) \leq f_{x,t-1}(\theta_{x,t-1}) - \beta t\|\theta_{x,t-1} - \theta\|^2$ for all $\theta \in M$. It then follows that for $\|y - x\| \leq \zeta$,

$$f_{y,t}(\theta_{x,t-1}) \leq f_{y,t}(\theta_{y,t}) = f_{x,t-1}(\theta_{y,t}) + \theta_{y,t}'(y-x) - \psi(\theta_{y,t})$$
$$\leq f_{x,t-1}(\theta_{x,t-1}) - \beta t\|\theta_{x,t-1} - \theta_{y,t}\|^2 + \theta_{y,t}'(y-x) - \psi(\theta_{y,t})$$
$$\leq f_{y,t}(\theta_{x,t-1}) - \beta t\|\theta_{x,t-1} - \theta_{y,t}\|^2 + (\zeta+\widetilde{\eta})\|\theta_{x,t-1} - \theta_{y,t}\|,$$

and therefore $\|\theta_{x,t-1} - \theta_{y,t}\| \leq (\zeta+\widetilde{\eta})/(\beta t)$. Hence (3.24) holds by setting $x = S_{t-1}$ and $y = S_t$. □

PROOF OF THEOREM 1. To simplify the notation, we will suppress the superscript $^{(1)}$ in $\widehat{\theta}_{t-1}^{(1)}$ below. By (2.4) and (3.20),

$$(3.27) \qquad h_{t-1}(\widetilde{\mathbf{Y}}_{t-1}^{(1)}) = \Big(\prod_{k=1}^{t-1} \bar{w}_k\Big)\exp[-\widehat{\theta}_{t-1}'\widetilde{S}_{t-1}^{(1)} + (t-1)\psi(\widehat{\theta}_{t-1})].$$

Making use of $E[f(\widetilde{\mathbf{Y}}_t^{(1)})|\mathcal{F}_{2t-2}] = f_{t-1}(\mathbf{Y}_{t-1}^{(1)})$, $E(\sup_{\theta \in M} e^{2\theta'\xi_t - 2\psi(\theta)}) < \infty$ and the independence of $\bar{w}_1^2 \cdots \bar{w}_{t-1}^2$ and ξ_t, we obtain from Lemma 1 and (3.27) that

$$(3.28) \qquad E\{[f_t(\widetilde{\mathbf{Y}}_t^{(1)}) - f_{t-1}(\mathbf{Y}_{t-1}^{(1)})]^2 h_{t-1}^2(\mathbf{Y}_{t-1}^{(1)})\}$$
$$\leq \quad E\{\bar{w}_1^2 \cdots \bar{w}_{t-1}^2 f_t^2(\widetilde{\mathbf{Y}}_t^{(1)}) / \exp[2\widehat{\theta}_{t-1}'S_{t-1}^{(1)} - 2(t-1)\psi(\widehat{\theta}_{t-1})]\}$$

$$\leq \quad e^{-2nI+o(n)}E(\bar{w}_1^2\cdots\bar{w}_{t-1}^2).$$

By (2.4) and Lemma 2,

$$E(\bar{w}_{t-1}^2|\mathcal{F}_{2(t-1)-2}) = \Big(m^{-1}\sum_{i=1}^{m}E[w_{t-1}(\widetilde{\mathbf{Y}}_{t-1}^{(i)})|S_{t-2}^{(i)}]\Big)^2$$

$$+m^{-2}\sum_{i=1}^{m}\mathrm{Var}[w_{t-1}(\widetilde{\mathbf{Y}}_{t-1}^{(i)})|S_{t-2}^{(i)}] \leq (1+Km^{-1})e^{2\epsilon_{t-1}},$$

and proceeding inductively yields

$$(3.29)\quad E(\bar{w}_1^2\cdots\bar{w}_{t-1}^2) \leq (1+Km^{-1})^{t-1}\exp\Big(\sum_{k=1}^{t-1}2\epsilon_k\Big) \leq e^{K(t-1)/m+o(n)}.$$

Similarly, under bootstrap or residual resampling,

$$(3.30)\qquad\qquad E[(\#_t^{(1)} - mw_t^{(1)})^2 f_t^2(\widetilde{\mathbf{Y}}_t^{(1)})h_t^2(\widetilde{\mathbf{Y}}_t^{(1)})]$$

$$= \quad m^{-1}\sum_{i=1}^{m}E[(\#_t^{(i)} - mw_t^{(i)})^2 f_t^2(\widetilde{\mathbf{Y}}_t^{(i)})h_t^2(\widetilde{\mathbf{Y}}_t^{(i)})]$$

$$\leq \quad e^{-2nI+o(n)}E(\bar{w}_1^2\cdots\bar{w}_t^2).$$

By (C1), $p_n = e^{-nI+o(n)}$ (see Theorem 2 of [6]) and hence it follows from (3.12) and (3.28)–(3.30) that both $\widehat{\alpha}_{\mathrm{R}}$ and $\widehat{\alpha}_{\mathrm{B}}$ are logarithmically efficient. □

The heuristic principle described in the paragraph following Example 1 can also be used to construct logarithmically efficient SISR procedures for Monte Carlo evaluation of (3.3), as illustrated in the following.

EXAMPLE 2. Let $T_c = \inf\{n : S_n \geq c\}$. Consider the estimation of $p_c = P\{T_c \leq n_1\}$ (i.e., with $d = 1$ and $g(x) = x$) when $\mu_0 < 0$ and $n_1 \sim ac$ for some $a > 1/\psi'(\theta_*)$, where θ_* is the unique positive root of $\psi(\theta_*) = 0$. We shall assume $2\theta_* \in \Theta$ and use the importance measure $\widetilde{Q} = P$ and resampling weights

$$w_t(\mathbf{Y}_t) = \begin{cases} e^{\theta_*\xi_t} & \text{if } t \leq T_c, \\ 1 & \text{if } n_1 > t > T_c. \end{cases}$$

Let $\eta(\mathbf{Y}_{T_c\wedge n_1}) = e^{\theta_*(S_{T_c\wedge n_1}-c)}$. Since $\eta(\mathbf{Y}_{T_c\wedge n_1}) \geq \mathbf{1}_{\{\max_{n\leq n_1}S_n\geq c\}}$, it follows that

$$(3.31)\quad f_t(\mathbf{Y}_t) = P\{\max_{n\leq n_1}S_n \geq c|\mathbf{Y}_t\} \leq E[\eta(\mathbf{Y}_{T_c\wedge n_1})|\mathbf{Y}_t] = e^{\theta_*(S_{T_c\wedge t}-c)}.$$

Making use of (3.31) in place of (3.11), we obtain that analogous to (3.13),
(3.32)
$$E\{[f_t(\widetilde{\mathbf{Y}}_t^{(1)}) - f_{t-1}(\mathbf{Y}_{t-1}^{(1)})]^2 h_{t-1}^2(\mathbf{Y}_{t-1}^{(1)})\} \le \Big(1 + \frac{K_*}{m}\Big)^{t-1} e^{-2\theta_* c} E(e^{2\theta_* \xi_t}),$$

where $K_* = E(e^{\theta_* \xi_1} - 1)^2$, and that analogous to (3.15),

$$(3.33) \quad E[(\#_t^{(1)} - mw_t^{(1)})^2 f_t^2(\widetilde{\mathbf{Y}}_t^{(1)}) h_t^2(\widetilde{\mathbf{Y}}_t^{(1)})] = O\Big(\Big(1 + \frac{K_*}{m}\Big)^{t-1} e^{-2\theta_* c}\Big).$$

Hence by (3.12) (with n_1 in place of n), (3.32) and (3.33),

$$m\text{Var}(\widehat{\alpha}_B) = O(n_1 \exp[(n_1 K_*/m) - 2\theta_* c]).$$

Since $n_1 = O(c)$ and $p_c/e^{-\theta_* c}$ is bounded away from 0 and ∞, as shown in [22], (3.7) also holds.

In Theorem 2 below, we provide the resampling weights for logarithmically efficient simulation of (3.3), for which the counterparts of (3.17) and (3.18) are also provided. The basic idea is to use the resampling weights (3.20) up to the stopping time

$$(3.34) \qquad\qquad T_c = \inf\{n \ge n_0 : ng(S_n/n) \ge c\} \wedge n_1.$$

THEOREM 2.   *Let $g(\mu_0) < a^{-1}$, $n_0 = \delta c + O(1)$ and $n_1 = ac + O(1)$ as $c \to \infty$ for some $a > \delta > 0$. Let $I = \inf\{\phi(\mu) : g(\mu) \ge \delta^{-1}\}$, and $M = \{\theta : \phi(\mu_\theta) \le I\}$. Let $\widetilde{Q} = P$ and assume that (C1)–(C2) hold for all $a^{-1} \le b \le \delta^{-1}$ and that*

(C3) $r := \sup_{\mu:g(\mu) \ge a^{-1}} \min\{g(\mu), \delta^{-1}\}/\phi(\mu) < \infty$.

Let $\widehat{\theta}_0 = 0$ and define for $1 \le t \le n_1 - 1$, $\widehat{\theta}_t = \arg\max_{\theta \in M}[\theta' S_t/t - \psi(\theta)]$ and

$$(3.35) \qquad w_t(\mathbf{Y}_t) = \begin{cases} e^{\widehat{\theta}_t' S_t - t\psi(\widehat{\theta}_t) - [\widehat{\theta}_{t-1}' S_{t-1} - (t-1)\psi(\widehat{\theta}_{t-1})]} & \text{if } t \le T_c \\ 1 & \text{if } n_1 > t > T_c. \end{cases}$$

Then (3.7) holds for $\widehat{\alpha}_B$ and with $\widehat{\alpha}_B$ replaced by $\widehat{\alpha}_R$ if $m \to \infty$ and $c \to \infty$.

PROOF. Let $u = (t-1) \wedge T_c^{(1)}$. By (2.4) and (3.35),

$$(3.36) \qquad h_{t-1}(\widetilde{\mathbf{Y}}_{t-1}^{(1)}) = \Big(\prod_{k=1}^{t-1} \bar{w}_k\Big) \exp[-(\widehat{\theta}_u^{(1)})' \widetilde{S}_u^{(1)} + u\psi(\widehat{\theta}_u^{(1)})].$$

Let $I_b = \inf\{\phi(\mu) : g(\mu) \geq b\}$. By Lemma 1,

$$(3.37) \quad f_t(\widetilde{\mathbf{Y}}_t^{(1)}) = P\{T_c^{(1)} \leq n_1 | \widetilde{\mathbf{Y}}_t^{(1)}\}$$

$$\leq \begin{cases} \sum_{n=t+1}^{n_1} e^{-nI_{c/n}+o(n)} \int_M e^{\theta'\tilde{S}_t^{(1)}-t\psi(\theta)} d\theta & \text{if } t < T_c^{(1)}, \\ 1 & \text{if } t \geq T_c^{(1)}. \end{cases}$$

Note that

$$\inf_{a^{-1} \leq b \leq \delta^{-1}} b^{-1} I_b = \min\left\{ \inf_{\mu:a^{-1}\leq g(\mu)\leq \delta^{-1}} \frac{\phi(\mu)}{g(\mu)}, \inf_{\mu:g(\mu)>\delta^{-1}} \frac{\phi(\mu)}{\delta^{-1}} \right\} = r^{-1},$$

by (C3). Hence by (3.36) and (3.37),

$$(3.38) \quad E\{[f_t(\widetilde{\mathbf{Y}}_t^{(1)}) - f_{t-1}(\mathbf{Y}_{t-1}^{(1)})]^2 h_{t-1}^2(\mathbf{Y}_{t-1}^{(1)})\} \leq e^{-2c/r+o(c)} E(\bar{w}_1^2 \cdots \bar{w}_{t-1}^2).$$

Similarly, it can be shown that under either bootstrap or residual resampling,

$$(3.39) \quad E[(\#_t^{(1)} - mw_t^{(1)})^2 f_t^2(\widetilde{\mathbf{Y}}_t^{(1)}) h_t^2(\widetilde{\mathbf{Y}}_t^{(1)})] \leq e^{-2c/r+o(c)} E(\bar{w}_1^2 \cdots \bar{w}_{t-1}^2).$$

By (C1) and Theorem 2 of [6], $p_c = e^{-c/r+o(c)}$ and hence it follows from (3.29), (3.38) and (3.39) that both $\widehat{\alpha}_{\mathrm{R}}$ and $\widehat{\alpha}_{\mathrm{B}}$ are logarithmically efficient. □

3.2. *Markovian extensions.* Let $\{(X_t, S_t) : t = 0, 1, \ldots,\}$ be a Markov additive process on $\mathcal{X} \times \mathbf{R}^d$ with transition kernel

$$\begin{aligned} P(x, A \times B) &:= P\{(X_1, S_1) \in A \times (B + s) | (X_0, S_0) = (x, s)\} \\ &= P\{(X_1, S_1) \in A \times B | (X_0, S_0) = (x, 0)\}. \end{aligned}$$

Let $\{X_n\}$ be aperiodic and irreducible with respect to some maximal irreducibility measure φ and assume that the transition kernel satisfies the minorization condition

$$(3.40) \qquad\qquad P(x, A \times B) \geq h(x, B)\nu(A)$$

for any measurable set $A \subset \mathcal{X}$, Borel set $B \subset \mathbf{R}^d$ and $s \in \mathbf{R}^d$ for some probability measure ν and measure $h(x, \cdot)$ that is positive for all x belonging to a φ-positive set. Ney and Nummelin [19] developed a theory to analyze large deviations properties of S_n under (3.40) or when its variant $P(x, A \times B) \geq h(x)\nu(A \times B)$ holds. Let τ be the first regeneration time and assume that $\Omega := \{(\theta, \zeta) : E_\nu e^{\theta'S_\tau - \tau\zeta} < \infty\}$ is an open neighborhood of 0. Then for all $\theta \in \Theta := \{\theta : (\theta, \zeta) \in \Omega \text{ for some } \zeta\}$, the kernel

$$(3.41) \qquad\qquad \widehat{P}_\theta(x, A) := \int e^{\theta's} P(x, A \times ds)$$

has a unique maximum eigenvalue $e^{\psi(\theta)}$, for which $\zeta = \psi(\theta)$ is the unique solution of the equation $E_\nu e^{\theta' S_\tau - \tau\zeta} = 1$, with corresponding right eigenfunctions $r(\cdot; \theta)$ and left eigenmeasures $\ell_\nu(\theta, \cdot)$ defined by

$$
\begin{aligned}
(3.42) \qquad r(x; \theta) &= E_x e^{\theta' S_\tau - \tau\psi(\theta)}, \\
\ell_x(\theta; A) &= E_x\Big(\sum_{n=0}^{\tau-1} e^{\theta' S_n - n\psi(\theta)} \mathbf{1}_{\{X_n \in A\}} \Big), \\
\ell_\nu(\theta; A) &= \int \ell_x(\theta; A) d\nu(x).
\end{aligned}
$$

Let π denote the stationary distribution of $\{X_n\}$ and let

$$
(3.43) \qquad \theta_\mu = (\nabla \psi)^{-1}(\mu).
$$

To begin with, consider the special case $d = 1$ and $g(x) = x$, for which the importance sampling measure with transition kernel

$$
(3.44) \qquad P_\theta(x, dy \times ds) := e^{\theta' s - \psi(\theta)}\{r(y; \theta)/r(x; \theta)\} P(x, dy \times ds)
$$

has been shown to be logarithmically efficient by Dupuis and Wang [16] and asymptotically optimal by Chan and Lai [8] for simulating the tail probability $P_{x_0}\{S_n/n \geq b\}$ when θ is chosen to be θ_b in (3.44). We shall show that by using SISR with $\widetilde{Q} = P$ and resampling weights $w_t(\mathbf{Y}_t) = e^{\theta_b \xi_t - \psi(\theta_b)}$, we can avoid computation of the eigenfunctions. To bring out the essence of the method, we first assume instead of the minorization condition (3.40) the stronger uniform recurrence condition

$$
(3.45) \qquad a_0 \nu(A \times B) \leq P(x, A \times B) \leq a_1 \nu(A \times B),
$$

for some $0 < a_0 < a_1$ and probability measure ν and for all $x \in \mathcal{X}$, measurable sets $A \subset \mathcal{X}$ and Borel sets $B \subset \mathbf{R}$. At the end of this section, we show how this assumption can be removed . Note that $\mathbf{Y}_t$ consists of (X_i, ξ_i), $i \leq t$, in the Markov case.

EXAMPLE 3. Let $b > E_\pi \xi_1$ and assume that $\theta_b \in \Theta$ and $E_\nu(e^{2\theta_b \xi_1 - 2\psi(\theta_b)}) < \infty$. We now extend Example 1 to Markov additive processes by showing that the choice $\widetilde{Q} = P$ and

$$
(3.46) \qquad w_t(\mathbf{Y}_t) = e^{\theta_b \xi_t - \psi(\theta_b)}
$$

results in logarithmically efficient simulation of $P_{x_0}\{S_n/n \geq b\}$. The dependence of the weights $w_t^{(i)}$ and $w_t^{(j)}$ for $i \neq j$, created from a combination

of the Markovian structure of the underlying process and bootstrap resampling, requires a more delicate peeling and induction argument than that in Example 1. By considering $\xi_t - \psi(\theta_b)/\theta_b$ instead of ξ_t, we may assume without loss of generality that $\psi(\theta_b) = 0$.

Let $\kappa = \sup_{x \in \mathcal{X}} r(x; \theta_b)/\inf_{x \in \mathcal{X}} r(x; \theta_b)$ and let E_θ be expectation with respect to P_θ. Then by (2.5) and (3.44),

$$
\begin{aligned}
f_t(\mathbf{Y}_t) &= P_{x_0}\{S_n/n \geq b|\mathbf{Y}_t\} = P\{S_n - S_t \geq nb - S_t|X_t, S_t\} \\
&= r(X_t; \theta_b)E_{\theta_b}[e^{-\theta_b(S_n - S_t)}\mathbf{1}_{\{S_n - S_t \geq nb - S_t\}}/r(X_n; \theta_b)|X_t, S_t] \\
&\leq \kappa e^{-\theta_b(nb - S_t)}.
\end{aligned}
$$

We shall show that

$$(3.47) \qquad E(\bar{w}_1^2 \cdots \bar{w}_t^2) = e^{o(t)} \text{ as } m \to \infty \text{ uniformly over } 1 \leq t \leq n - 1.$$

Then logarithmic efficiency of bootstrap resampling follows from (3.12)–(3.15). We first show that for any $k < t$ and $i \neq j$,

$$
(3.48) \qquad E\{\bar{w}_k^2(E_{X_k^{(i)}}e^{\theta_b S_{t-k}})(E_{X_k^{(j)}}e^{\theta_b S_{t-k}})|\mathcal{F}_{2k-2}\}
$$
$$
\leq m^{-2}\sum_{u \neq v}(E_{X_{k-1}^{(u)}}e^{\theta_b S_{t-k+1}})(E_{X_{k-1}^{(v)}}e^{\theta_b S_{t-k+1}}) + m^{-1}\beta,
$$

where $\beta = \sup_{h \geq 0, x \in \mathcal{X}} E_x\{e^{2\theta_b \xi_1}(E_{X_1}e^{\theta_b S_h})^2\}$, which is finite by (3.45). Note that $\bar{w}_k$ is measurable with respect to $\mathcal{F}_{2k-1}$ and that under bootstrap resampling, $X_k^{(i)}$ and $X_k^{(j)}$ are independent conditioned on $\mathcal{F}_{2k-1}$. Moreover, since $X_k^{(1)} = \widetilde{X}_k^{(\ell)}$ with probability $w_k^{(\ell)} = w_k(\widetilde{\mathbf{Y}}_k^{(\ell)})/\sum_{j=1}^m w_k(\widetilde{\mathbf{Y}}_k^{(j)})$,

$$
E\{\bar{w}_k(E_{X_k^{(1)}}e^{\theta_b S_{t-k}})|\mathcal{F}_{2k-1}\} = \bar{w}_k \sum_{u=1}^m w_k^{(u)} E_{\widetilde{X}_k^{(u)}}e^{\theta_b S_{t-k}},
$$

which is equal to $m^{-1}\sum_{u=1}^m e^{\theta_b \tilde{\xi}_k^{(u)}} E_{\widetilde{X}_k^{(u)}}e^{\theta_b S_{t-k}}$ in view of (3.46) and that $\psi(\theta_b) = 0$. Hence

$$
(3.49) \qquad E\{\bar{w}_k^2(E_{X_k^{(i)}}e^{\theta_b S_{t-k}})(E_{X_k^{(j)}}e^{\theta_b S_{t-k}})|\mathcal{F}_{2k-1}\}
$$
$$
= \left(m^{-1}\sum_{u=1}^m e^{\theta_b \tilde{\xi}_k^{(u)}} E_{\widetilde{X}_k^{(u)}}e^{\theta_b S_{t-k}}\right)^2
$$
$$
= m^{-2}\sum_{u \neq v}(e^{\theta_b \tilde{\xi}_k^{(u)}} E_{\widetilde{X}_k^{(u)}}e^{\theta_b S_{t-k}})(e^{\theta_b \tilde{\xi}_k^{(v)}} E_{\widetilde{X}_k^{(v)}}e^{\theta_b S_{t-k}})
$$
$$
+ m^{-2}\sum_{u=1}^m e^{2\theta_b \tilde{\xi}_k^{(u)}} (E_{\widetilde{X}_k^{(u)}}e^{\theta_b S_{t-k}})^2.
$$

Since $(\widetilde{\xi}_k^{(u)}, \widetilde{X}_k^{(u)})$ and $(\widetilde{\xi}_k^{(v)}, \widetilde{X}_k^{(v)})$ are independent conditioned on $\mathcal{F}_{2k-2}$ for $u \neq v$ and $E[e^{\theta_b \widetilde{\xi}_k^{(i)}} (E_{\widetilde{X}_k^{(i)}} e^{\theta_b S_{t-k}}) | \mathcal{F}_{2k-2}] = E_{X_{k-1}^{(i)}} e^{\theta_b S_{t-k+1}}$, (3.48) follows from (3.49).

We shall show using (3.48) and induction, that

$$(3.50) \quad E(\bar{w}_1^2 \cdots \bar{w}_k^2) \leq \gamma^2 (1 + m^{-1}\beta)^k \text{ where } \gamma = \sup_{x \in \mathcal{X}, h \geq 0} E_x e^{\theta_b S_h} (\geq 1).$$

For $k = 1$,

$$E\bar{w}_1^2 = m^{-2} \sum_{i \neq j} E_{x_0} e^{\theta_b \xi_1^{(i)}} E_{x_0} e^{\theta_b \xi_1^{(j)}} + m^{-2} \sum_{i=1}^{m} E_{x_0} e^{2\theta_b \xi_1^{(i)}} \leq \gamma^2 + m^{-1}\beta$$

and indeed (3.50) holds. If (3.50) holds for all $k < t$, then by repeated application of (3.48), starting from $k = t$, we obtain

$$
\begin{aligned}
E(\bar{w}_1^2 \cdots \bar{w}_t^2) &\leq (E_{x_0} e^{\theta_b S_t})^2 + m^{-1}\beta \sum_{k=0}^{t-1} E(\bar{w}_1^2 \cdots \bar{w}_k^2) \\
&\leq \gamma^2 \Big\{ 1 + m^{-1}\beta \sum_{k=0}^{t-1} (1 + m^{-1}\beta)^k \Big\} = \gamma^2 (1 + m^{-1}\beta)^t,
\end{aligned}
$$

and (3.50) indeed holds for $k = t$. Hence (3.47) is true and logarithmic efficiency is attained.

The peeling argument used to derive (3.48) and (3.50) can also be used to extend Theorems 1 and 2, which hold for general g, to the following.

THEOREM 3. (a) *Let* M, $\widehat{\theta}_t$ *and* $w_t(\mathbf{Y}_t)$ *be the same as in Theorem 1. Then Theorem 1 still holds when the i.i.d. assumption on* ξ_t *is replaced by the uniform recurrence condition* (3.45) *on the Markov additive process* $(X_t, S_t = \xi_1 + \cdots + \xi_t)$ *and assumption* (C2) *is generalized to*

$$(3.51) \qquad \int_{\mathbf{R}^d} e^{2\kappa\|\xi\|} \nu(\mathcal{X}, d\xi) < \infty, \text{ where } \kappa = \sup_{\theta \in M} \|\theta\|.$$

(b) *Let* M, $\widehat{\theta}_t$ *and* $w_t(\mathbf{Y}_t)$ *be the same as in Theorem 2. Then Theorem 2 still holds when the i.i.d. assumption on* ξ_t *is replaced by the uniform recurrence condition* (3.45) *and assumption* (C2) *is generalized to* (3.51).

Note that $\widetilde{Q} = P$ in Theorem 3. We next show how the uniform recurrence assumption (3.45) can be removed, extending the preceding results on the

logarithmic efficiency of suitably chosen SISR procedures to more general Markov additive processes such that for some $\theta \in \Theta$, $0 < \beta < 1$, function $u : \mathcal{X} \to [1, \infty)$ and measurable set C,

(U1) $\sup_{x \in C} u(x) < \infty, \quad \int_{\mathcal{X}} u(x) d\nu(x) < \infty,$

$\quad\quad \sup_{x \in C} \ell_x(\theta; C) < \infty, \quad \int_{\mathcal{X}} \ell_x(\theta; C) d\nu(x) < \infty,$

(U2) $E_x\{e^{\theta' \xi_1 - \psi(\theta)} u(X_1)\} \le (1 - \beta) u(x)$ for $x \notin C,$

(U3) $a := \sup_{x \in C} E_x\{e^{\theta' \xi_1 - \psi(\theta)} u(X_1)\} < \infty,$

(U4) $K_1 := \sup_{x \in \mathcal{X}} E_x\{e^{2\theta' \xi_1 - 2\psi(\theta)} u^2(X_1)/u^2(x)\} < \infty.$

We illustrate in Section 4, Example 5 how (U1)–(U4) can be checked in a concrete example. Condition (U1), in which ℓ_x is defined in (3.42), holds when C is bounded and ν has support on a compact set. Conditions (U2)–(U4) are often called "drift conditions", see [8]. Although the arguments are essentially modifications of the peeling idea in Example 3 by making use of (U1)–(U4), they are considerably more complicated than those in the uniformly recurrent case. We therefore only consider the univariate linear case ($d = 1$, $g(y) = y$) in the following theorem to indicate the basic ideas without getting into the details of these modifications, such as replacing for general g the θ_b in (3.52) by sequential estimates $\widehat{\theta}_t$, as in (3.20) and (3.35).

THEOREM 4.   *Let $b > E_\pi \xi_1$ and assume that (U1)–(U4) hold for $\theta = \theta_b$. Let $\widetilde{Q} = P$ and*

(3.52) $$w_t(\mathbf{Y}_t) = e^{\theta_b \xi_t - \psi(\theta_b)} u(X_t)/u(X_{t-1}).$$

Then (3.6) holds with $p_n = P_{x_0}\{S_n/n \ge b\}$, for $\widehat{\alpha}_{\mathrm{B}}$ or $\widehat{\alpha}_{\mathrm{R}}$, as $n \to \infty$ and $m \to \infty$.

PROOF. By considering $\xi_t - \psi(\theta_b)/\theta_b$ instead of ξ_t, we assume without loss of generality that $\psi(\theta_b) = 0$. By (2.4) and (3.52),

(3.53) $$h_{t-1}(\widetilde{\mathbf{Y}}_{t-1}^{(1)}) = \Big(\prod_{k=1}^{t-1} \bar{w}_k\Big) e^{-\theta_b \widetilde{S}_{t-1}^{(1)}} u(x_0)/u(\widetilde{X}_{t-1}^{(1)}).$$

It will be shown in the Appendix that

(3.54) $$K_2 := \sup_{x \in \mathcal{X}, h \ge 0} E_x\{e^{\theta_b S_h} u(X_h)/u(x)\} < \infty.$$

Note that

(3.55) $$\begin{aligned} f_t(\mathbf{Y}_t) &= E_{x_0}(\mathbf{1}_{\{S_n/n \ge b\}}|\mathbf{Y}_t) \le e^{-\theta_b n b} E_{x_0}(e^{\theta_b S_n}|\mathbf{Y}_t) \\ &= e^{\theta_b(S_t - nb)} E_{X_t}(e^{\theta_b S_{n-t}}) \le K_2 e^{\theta_b(S_t - nb)} u(X_t). \end{aligned}$$

Since $E_{x_0}[f_t(\widetilde{\mathbf{Y}}_t^{(1)})|\mathcal{F}_{2t-2}] = f_{t-1}(\mathbf{Y}_{t-1}^{(1)})$, it follows from (3.53), (3.55) and (U3) that

$$(3.56) \qquad E_{x_0}\{[f_t(\widetilde{\mathbf{Y}}_t^{(1)}) - f_{t-1}(\mathbf{Y}_{t-1}^{(1)})]^2 h_{t-1}^2(\mathbf{Y}_{t-1}^{(1)})\}$$
$$\leq K_2^2 e^{-2n\theta_b b} E_{x_0}\{(\bar{w}_1 \cdots \bar{w}_{t-1})^2 e^{2\theta_b \tilde{\xi}_t^{(1)}} u^2(x_0) u^2(\widetilde{X}_t^{(1)})/u^2(X_{t-1}^{(1)})\}$$
$$\leq \beta e^{-2n\theta_b b} E_{x_0}(\bar{w}_1^2 \cdots \bar{w}_{t-1}^2),$$

where $\beta = K_1 K_2^2 u^2(x_0)$.

By (3.53) and (3.55), under either bootstrap or residual resampling,

$$(3.57) \qquad E_{x_0}[(\#_t^{(1)} - mw_t^{(1)})^2 f_t^2(\widetilde{\mathbf{Y}}_t^{(1)}) h_t^2(\widetilde{\mathbf{Y}}_t^{(1)})]$$
$$= m^{-1}\sum_{i=1}^m E_{x_0}[(\#_t^{(i)} - mw_t^{(i)})^2 f_t^2(\widetilde{\mathbf{Y}}_t^{(i)}) h_t^2(\widetilde{\mathbf{Y}}_t^{(i)})]$$
$$\leq K_2^2 E_{x_0}(\bar{w}_1^2 \cdots \bar{w}_t^2) e^{-2n\theta_b b} u^2(x_0).$$

In view of (3.12), it now remains to show (3.47). It follows from the proof of (3.48) that for any $k < t$ and $i \neq j$,

$$E_{x_0}\Big\{\bar{w}_k^2\Big(\frac{E_{X_k^{(i)}}[e^{\theta_b S_{t-k}} u(X_{t-k})]}{u(X_k^{(i)})}\Big)\Big(\frac{E_{X_k^{(j)}}[e^{\theta_b S_{t-k}} u(X_{t-k})]}{u(X_k^{(j)})}\Big)\Big|\mathcal{F}_{2k-2}\Big\}$$
$$\leq m^{-2}\sum_{v \neq w}\Big(\frac{E_{X_{k-1}^{(v)}}[e^{\theta_b S_{t-k+1}} u(X_{t-k+1})]}{u(X_{k-1}^{(v)})}\Big)\Big(\frac{E_{X_{k-1}^{(w)}}[e^{\theta_b S_{t-k+1}} u(X_{t-k+1})]}{u(X_{k-1}^{(w)})}\Big)$$
$$+ m^{-1}\beta.$$

An argument similar to that in (3.48) and (3.50) can be used to show that

$$E_{x_0}(\bar{w}_1^2 \cdots \bar{w}_k^2) \leq K_2^2(1 + m^{-1}\beta)^k.$$

Hence (3.47) again holds and (3.6) follows from (3.56) and (3.57). $\qquad \square$

3.3. *Implementation, estimation of standard errors and discussion.* As explained in the first paragraph of Section 3.1, at every stage t, the SISR procedure carries out importance sampling sequentially within each simulated trajectory, but performs resampling across the m trajectories. Since the computation time for resampling increases with m, it is more efficient to divide the m trajectories into r subgroups of size k so that $m = kr$ and resampling is performed within each subgroup of k trajectories, independently of the other subgroups. This method also has the advantage of

providing a direct estimate of the standard error of the Monte Carlo estimate $\bar{\alpha} := r^{-1} \sum_{i=1}^{r} \widehat{\alpha}_i$, where $\widehat{\alpha}_i$ denotes the SISR estimate of α (using either bootstrap or residual resampling) based on the ith subgroup of simulated trajectories. Specifically, we can estimate the standard error of $\bar{\alpha}$ by $\widehat{\sigma}/\sqrt{r}$, where

$$(3.58) \qquad \widehat{\sigma}^2 = (r-1)^{-1} \sum_{i=1}^{r} (\widehat{\alpha}_i - \bar{\alpha})^2.$$

In Section 2 we have considered the case of fixed n as $m \to \infty$ and provided estimates of the standard errors of the asymptotically normal $\widehat{\alpha}_{\mathrm{B}}$ and $\widehat{\alpha}_{\mathrm{R}}$. The validity of these estimates is unclear for the case $n \to \infty$ and $m \to \infty$, as considered in this section that involves large deviations theory instead of central limit theorems. By choosing $m = kr$ with $k \to \infty$ and $r \to \infty$ in (3.58), we still have a consistent estimate $\widehat{\sigma}/\sqrt{r}$ of the standard error in the large deviations setting with $n \to \infty$.

The resampling weights in Theorems 1 and 2 have closed-form expressions in terms of the cumulant generating function $\psi(\theta)$ in the i.i.d. case or the logarithm $\psi(\theta)$ of the largest eigenvalue of the kernel (3.41) in the Markov case. When $\psi(\theta)$ does not have explicit formulas, we can use numerical approximations and thereby approximate the logarithmically efficient resampling weights, as will be illustrated in Example 5 below. This is, therefore, much more flexible than logarithmically efficient importance sampling which involves sampling from the efficient importance measure that involves both the eigenvalue and corresponding eigenfunction in the Markov case; see [5, 8, 10, 16, 21]. Note that approximating the eigenvalue and eigenfunction usually does not result in an importance (probability) measure and therefore requires an additional task of computing the normalizing constants.

The basic ideas in Examples 1 and 2 and Sections 3.1 and 3.2 can be extended to more general rare events of the form $\{\mathbf{X}_T \in \Gamma\}$ and more general stochastic sequences $\mathbf{X}_t$ and stopping times T. To evaluate $P\{\mathbf{X}_T \in \Gamma\}$ by Monte Carlo, it would be ideal to sample from the importance measure Q for which

$$(3.59) \qquad \frac{dQ}{dP}(\mathbf{X}_t) = P\{\mathbf{X}_T \in \Gamma | \mathbf{X}_t\}/P\{\mathbf{X}_T \in \Gamma\} \text{ for } t \le T,$$

because the corresponding Monte Carlo estimate of $P\{\mathbf{X}_T \in \Gamma\}$ would have variance 0; see [16, p.2]. This is clearly not feasible because the right-hand side of (3.59) involves the conditional probabilities $P\{\mathbf{X}_T \in \Gamma | \mathbf{X}_t\}$ and its expectation $P\{\mathbf{X}_T \in \Gamma\}$, which is an unknown quantity to be determined. On the other hand, SISR enables one to ignore the normalizing factor $P\{\mathbf{X}_T \in \Gamma\}$ and to use tractable approximations to $P\{\mathbf{X}_T \in \Gamma | \mathbf{X}_t\}$,

as in Example 1, in coming up with a logarithmically efficient Monte Carlo estimate of $P\{\mathbf{X}_T \in \Gamma\}$.

4. Illustrative examples.   We use the following two examples to illustrate Theorems 1 and 4.

EXAMPLE 4.   Let $X_1, X_2, \ldots$ be i.i.d. random variables with $EX_1 = 0$. Let $\xi_i = (X_i, X_i^2)$ and $S_n = \xi_1 + \cdots + \xi_n$. Define $g(y, v) = y/\sqrt{v}$ for $y \in \mathbf{R}$ and $v > 0$, and note that $g(S_n/n)$ is the self-normalized sum of the X_i's. There is an extensive literature on the large deviation probability $p_n = P\{g(S_n/n) \geq b\}$; see [12]. Consider the case $b = 1/\sqrt{2}$ and X_i having the density function

$$f(x) = \frac{1}{2\sqrt{2\pi}}(e^{-(x-1)^2/2} + e^{-(x+1)^2/2}), \quad x \in \mathbf{R},$$

with respect to the Lebesgue measure. Thus, X_i is a mixture of $N(1,1)$ and $N(-1,1)$. In this case, $\Theta = \{(\theta_1, \theta_2) : \theta_2 < 1/2\}$, $\Lambda = \{(y, v) : v \geq y^2\}$ and

$$\log(Ee^{\theta_1 X_1 + \theta_2 X_1^2}) = \log\left(\frac{1}{2}\right) + \frac{1}{2} - \frac{\theta_1^2 + 1}{2 - 4\theta_2} + \log\left(\frac{e^{\theta_1/(1-2\theta_2)} + e^{-\theta_1/(1-2\theta_2)}}{\sqrt{1 - 2\theta_2}}\right),$$

for $\theta \in \Theta$. The infimum of the rate function over the one-dimensional manifold $N = \{(y, v) : y = \sqrt{v/2}\}$ is $I = .324$ and is attained at $(y, v) = (1, 2)$. Then $M = \{\theta = (\theta_1, \theta_2) : \phi(y_\theta, v_\theta) \leq I\}$, see (3.18) and Theorem 1. We implement SISR with bootstrap resampling as described in Section 3.3, with $m = 10{,}000$ particles, divided into 100 groups each having 100 particles. The results, in the form of mean±standard error and for $n = 15, 20$ and 25, are summarized in Table 1, which also compares them to corresponding results obtained by direct Monte Carlo with $m = 10{,}000$ in (2.1) and (2.2). Table 1 shows 18-fold variance reduction by using SISR when $n = 15$, 25-fold variance reduction when $n = 20$ and that direct Monte Carlo fails when $n = 25$.

<div align="center">INSERT TABLE 1 ABOUT HERE</div>

EXAMPLE 5.   Let $\zeta_1, \zeta_2, \ldots, \gamma_1, \gamma_2, \ldots$ be i.i.d. standard normal random variables and let

(4.1) $X_{n+1} = \lambda(X_n) + \zeta_{n+1}, \qquad \xi_n = X_n + \gamma_n,$

where $\lambda(x)$ is a monotone increasing, piecewise linear function given by

$$\lambda(x) = x\mathbf{1}_{\{|x|\leq 1\}} + \left(\frac{x+1}{2}\right)\mathbf{1}_{\{x>1\}} + \left(\frac{x-1}{2}\right)\mathbf{1}_{\{x<-1\}}.$$

Let $\theta > 0$. We now show that (U1)–(U4) hold for $u(x) = e^{2.1\theta x^+}$ and $C = (-\infty, \rho]$, where $\rho \geq 1$ is chosen large enough so that (U2) holds, as shown below. Since $(a + b)^+ \leq a + b^+$ for $a > 0$ and since $e^{2.05\theta x} \leq e^{-.05\theta x}u(x)$, it follows that for $x > \rho$,

$$
\begin{aligned}
E_x\{e^{\theta\xi_1 - \psi(\theta)}u(X_1)\} &= E\{e^{\theta x + \theta\gamma_1 - \psi(\theta) + 2.1\theta((\frac{x+1}{2}) + \zeta_1)^+}\} \\
&\leq u(x)e^{-.05\theta x}E\{e^{\theta\gamma_1 - \psi(\theta) + 1.05\theta + 2.1\theta\zeta_1^+}\},
\end{aligned}
$$

and therefore (U2) holds if ρ is large enough. It is easy to check that (U3) holds. Note that $\sup_{x \in (-\infty, 1]} E_x[e^{2\theta\xi_1 - 2\psi(\theta)}u^2(X_1)] < \infty$ and that for $x > 1$,

$$
\begin{aligned}
&E_x[e^{2\theta\xi_1 - 2\psi(\theta)}u^2(X_1)]/u^2(x) \\
&= E[e^{2\theta x + 2\theta\gamma_1 - 2\psi(\theta) + 4.2\theta((\frac{x+1}{2}) + \zeta_1)^+}]/e^{4.2\theta x^+} \\
&\leq (e^{-.1\theta x} \wedge e^{2\theta x})E[e^{2\theta\gamma_1 - 2\psi(\theta) + 2.1\theta + 4.2\theta\zeta_1^+}] \to 0 \text{ as } x \to \infty,
\end{aligned}
$$

and therefore (U4) holds. Since $\lim_{x \to -\infty} E_x(e^{\theta\xi_1 - \psi(\theta)}) = 0$, it follows that $\lim_{x \to -\infty} \ell_x(\theta; C) = 0$; moreover $u(x) = 1$ for all $x \leq 0$ and hence (U1) also holds.

We compute $P_0\{S_n/n \geq 2.5\}$ for SISR using resampling, with $m = 10,000$ particles divided into 100 groups, each having 100 particles, and with resampling weights (3.52) for which the following procedure is used to provide a numerical approximation for $\theta_{2.5}$. First note that by (4.1),

$$
(4.2) \qquad\qquad E_x e^{\theta\xi_1} = e^{\theta^2/2} E_x e^{\theta X_1}.
$$

The procedure involves a finite-state Markov chain approximation to (4.1) with states x_i and transition probabilities p_{ij} $(1 \leq i, j \leq 1000)$ given by

$$
x_i = \frac{i}{100} - 2.505, \quad p_{ij} = e^{-(x_j - \lambda(x_i))^2/2} \Big/ \sum_{k=1}^{1000} e^{-(x_k - \lambda(x_i))^2/2}.
$$

For given θ, it approximates $\psi(\theta)$ by $\theta^2/2 + \widetilde{\psi}(\theta)$, where $e^{\widetilde{\psi}(\theta)}$ is the largest eigenvalue of the matrix $(e^{\theta x_j}p_{ij})_{1 \leq i,j \leq 1000}$, in view of (3.41) and (4.2). Since $\psi'(\theta_{2.5}) = 2.5$ by (3.43), it uses Brent's method [20] that involves bracketing followed by efficient search to find the positive root $\widetilde{\theta}_{2.5}$ of the equation $\widetilde{\psi}(\theta) + \theta^2/2 = 2.5\theta$, noting that $\widetilde{\psi}(0) = 0$. The root $\widetilde{\theta}_{2.5} = .273$ is then used as an approximation to $\theta_{2.5}$ in (3.52). Table 2 gives the results, in the form of mean±standard error, for the SISR (with several choices of θ in (3.52), including $\theta = \widetilde{\theta}_{2.5}$) and direct Monte Carlo estimates of $P_0\{S_n/n \geq 2.5\}$. It shows a variance reduction of 35 times for $n = 15$ and 80 times for $n = 20$ over direct Monte Carlo when $\widetilde{\theta}_{2.5}$ is used as an approximation to $\theta_{2.5}$ in the resampling weights (3.52) for SISR. When $n = 25$, direct Monte Carlo fails while the SISR estimate still has a reasonably small standard error.

APPENDIX A: PROOF OF (3.19) AND (3.54)

PROOF OF (3.19). For $0 < \epsilon < I$, let

$$M_\epsilon = \{\theta : \phi(\mu_\theta) = I - \epsilon\}, \quad H(\theta) = \{\mu \in \Lambda^o : \theta'(\mu - \mu_\theta) \geq 0\}.$$

If $\mu \in H(\theta)$, then $\theta'\mu \geq \theta'\mu_\theta$ and therefore

$$(A.1) \qquad \phi(\mu) = \sup_{\widetilde{\theta}}\{\widetilde{\theta}'\mu - \psi(\widetilde{\theta})\} \geq \theta'\mu - \psi(\theta) \geq \theta'\mu_\theta - \psi(\theta) = I - \epsilon.$$

Moreover, for $\theta \in M_\epsilon$, $H(\theta)$ is a closed half-space whose boundary is the tangent space of $\{\mu : \phi(\mu) = I - \epsilon\}$ at μ_θ. Hence

$$(A.2) \qquad \phi(\mu) \neq I - \epsilon \text{ for } \mu \in \Lambda^o \setminus \bigcup_{\theta \in M_\epsilon} H(\theta).$$

Making use of this and (A.1), we next show that

$$(A.3) \qquad \bigcup_{\theta \in M_\epsilon} H(\theta) = \{\mu : \phi(\mu) \geq I - \epsilon\},$$

and therefore by (3.17),

$$(A.4) \qquad \Gamma := \{\mu : g(\mu) \geq b\} \subset \{\mu : \phi(\mu) \geq I - \epsilon\} = \bigcup_{\theta \in M_\epsilon} H(\theta).$$

By (A.1), $\bigcup_{\theta \in M_\epsilon} H(\theta) \subset \{\mu : \phi(\mu) \geq I - \epsilon\}$. Therefore it suffices for the proof of (A.3) to show that $\{\mu : \phi(\mu) < I - \epsilon\} \supset \Lambda^o \setminus \bigcup_{\theta \in M_\epsilon} H(\theta)$. Suppose this is not the case. Then there exists $\mu_1 \in \Lambda^o \setminus \bigcup_{\theta \in M_\epsilon} H(\theta)$ such that $\phi(\mu_1) \geq I - \epsilon$. Since $\Lambda^o \setminus \bigcup_{\theta \in M_\epsilon} H(\theta) \supset \{\mu : \phi(\mu) < I - \epsilon\}$, there exists $\mu_2 \in \Lambda^o \setminus \bigcup_{\theta \in M_\epsilon} H(\theta)$ such that $\phi(\mu_2) < I - \epsilon$. By continuity of ϕ, there exists $\rho \in (0, 1)$ such that $\phi(\rho\mu_1 + (1 - \rho)\mu_2) = I - \epsilon$. Since $\Lambda^o \setminus H(\theta)$ is a half-space, $\Lambda^o \setminus \bigcup_{\theta \in M_\epsilon} H(\theta) = \bigcap_{\theta \in M_\epsilon}(\Lambda^o \setminus H(\theta))$ is convex and therefore $\rho\mu_1 + (1 - \rho)\mu_2 \in \Lambda^o \setminus \bigcup_{\theta \in M_\epsilon} H(\theta)$, but this contradicts (A.2), thereby proving (A.3).

Define the measure Q by

$$\frac{dQ}{dP}(\mathbf{Y}_n) = \int_M e^{\theta' S_n - n\psi(\theta)} d\theta \Big/ \mathrm{vol}(M),$$

where $\mathrm{vol}(M)$ is the volume of M. Let $\mu_n = S_n/n$ and $h_n(\theta) = \theta'\mu_n - \psi(\theta)$. From (A.4), it follows that if $\mu_n \in \Gamma$ then there exists $\theta_* \in M_\epsilon$ such that $\theta'_*(\mu_n - \mu_{\theta_*}) \geq 0$, and therefore

$$(A.5) \qquad h_n(\theta_*) = \theta'_*\mu_n - \psi(\theta_*) \geq \theta'_*\mu_{\theta_*} - \psi(\theta_*) = \phi(\mu_{\theta_*}) = I - \epsilon,$$

since $\theta_* \in M_\epsilon$. Let $B_n = \{\theta : (\theta - \theta_*)'\nabla h_n(\theta_*) \geq 0, \; \|\theta - \theta_*\| \leq n^{-1/2}\}$. Then for all $\theta \in B_n$, $h_n(\theta) = h_n(\theta^*) + (\theta - \theta^*)'\nabla h_n(\theta_*) - (\theta - \theta_*)'\nabla^2\psi(\theta_*)(\theta - \theta_*)/2 + o(\|\theta - \theta_*\|^2)$, and therefore by (A.5) and the definition of B_n,

$$h_n(\theta) \geq I - \epsilon - (K+1)/(2n) \text{ for all large } n,$$

where $K = \sup_{\theta \in M}\|\nabla^2\psi(\theta)\|$. Hence for all large n,

$$(A.6) \qquad \frac{dQ}{dP}(\mathbf{Y}_n) \;\geq\; \mathbf{1}_{\{\mu_n \in \Gamma\}}\int_{B_n}\exp\{nh_n(\theta)\}d\theta\Big/\mathrm{vol}(M)$$

$$\geq\; \mathbf{1}_{\{\mu_n \in \Gamma\}}(c_d/2)e^{nI - n\epsilon - (K+1)/2}n^{-d/2}/\mathrm{vol}(M),$$

in which c_d denotes the volume of the d-dimensional unit ball. Letting $\epsilon \to 0$ in (A.6) yields $(dQ/dP)(\mathbf{Y}_n) \geq e^{nI + o(n)}\mathbf{1}_{\{\mu_n \in \Gamma\}}$, in which $o(n)$ is uniform in $\mathbf{Y}_n$. Hence

$$P\{g(S_n/n) \geq b|\mathbf{Y}_k\} \;=\; E_Q\Big[\frac{dP}{dQ}(\mathbf{Y}_n)\mathbf{1}_{\{S_n/n \in \Gamma\}}\frac{dQ}{dP}(\mathbf{Y}_k)\Big|\mathbf{Y}_k\Big]$$

$$\leq\; e^{-nI + o(n)}\frac{dQ}{dP}(\mathbf{Y}_k),$$

proving (3.19). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

To prove (3.54), we use ideas similar to those in the proof of Lemma 1 of [7] and the following result of [19, p.568].

LEMMA 3.   Let $\tau(0) = 0$. Under (3.40), there exist regeneration times $\tau(i)$, $i \geq 1$, such that

(i) $\tau(i+1) - \tau(i)$, $i \geq 0$, are i.i.d. random variables,

(ii) $\{X_{\tau(i)}, \ldots, X_{\tau(i+1)-1}, \xi_{\tau(i)+1}, \ldots, \xi_{\tau(i+1)}\}$, $i = 0, 1, \ldots$ are independent blocks,

(iii) $X_{\tau(i)}$ has distribution ν for all $i \geq 1$.

PROOF OF (3.54).   Let $\widetilde{\ell}_x = E_x\{\sum_{n=1}^{\tau}e^{\theta_b S_n}u(X_n)\}$, $\widetilde{\ell}_\nu = \int\widetilde{\ell}_x d\nu(x)$ and $A = \{\tau(i) : i \geq 1\}$. Since $u \geq 1$,

$$(A.7) \qquad E_x\{e^{\theta_b S_k}u(X_k)\} = E_x\{e^{\theta_b S_k}u(X_k)\mathbf{1}_{\{\tau \geq k\}}\}$$

$$+ \sum_{j=1}^{k-1}E_x(e^{\theta_b S_j}\mathbf{1}_{\{j \in A\}})E_\nu(e^{\theta_b S_{k-j}}u(X_{k-j})\mathbf{1}_{\{\tau \geq k-j\}})$$

$$\leq \widetilde{\ell}_x + \widetilde{\ell}_\nu[\sup_{j \geq 1}E_x(e^{\theta_b S_j}\mathbf{1}_{\{j \in A\}})].$$

Let $0 < \sigma = \sigma(1) < \sigma(2) < \cdots$ be the hitting times of C. Then
(A.8)

$$\widetilde{\ell}_x \leq E_x\Big\{ \sum_{n=1}^{\sigma} e^{\theta_b S_n} u(X_n)\Big\} + E_x\Big\{ \sum_{k:\sigma(k)<\tau} e^{\theta_b S_{\sigma(k)}} \sum_{n=\sigma(k)+1}^{\sigma(k+1)} e^{\theta_b(S_n - S_{\sigma(k)})} u(X_n)\Big\}.$$

Let $y \in \mathcal{X}$. By (U2), for all $n \geq 2$,

$$E_y\{e^{\theta_b S_n} u(X_n)\mathbf{1}_{\{n \leq \sigma\}}\} \leq (1-\beta)E_y(e^{\theta_b S_{n-1}} u(X_{n-1})\mathbf{1}_{\{n-1 \leq \sigma\}}),$$

from which it follows by proceeding inductively and applying (U3) that

$$(A.9) \qquad E_y\Big\{ \sum_{n=1}^{\sigma} e^{\theta_b S_n} u(X_n)\Big\} \leq \beta^{-1} \max\{a, (1-\beta)u(y)\} \leq \alpha u(y),$$

where $\alpha = \beta^{-1} \max\{a, (1-\beta)\}$. Substitution of (A.9) into (A.8) then yields

$$(A.10) \ \widetilde{\ell}_x \leq \alpha\Big\{u(x) + E_x\Big(\sum_{n=0}^{\tau-1} e^{\theta_b S_n} u(X_n)\mathbf{1}_{\{X_n \in C\}}\Big)\Big\} \leq \alpha u(x) + \eta \ell_x(\theta_b; C),$$

where $\eta = \sup_{y \in C} u(y)$. Since $\int_{\mathcal{X}} u(x)d\nu(x) < \infty$ and $\int_{\mathcal{X}} \ell_x(\theta_b; C)d\nu(x) < \infty$, it follows from (A.10) that $\widetilde{\ell}_\nu < \infty$. Combining

$$\ell_x(\theta_b; C) \leq E_x(e^{\theta_b S_\sigma})[\sup_{y \in C} \ell_y(\theta_b; C)]$$

with (A.9) yields

$$(A.11) \qquad\qquad \sup_{x \in \mathcal{X}}\{\ell_x(\theta_b; C)/u(x)\} < \infty.$$

Let Q^* be a probability measure under which

$$\frac{dQ^*}{dP_\nu}(\{(X_t, S_t) : t \leq \tau(i)\}) = e^{\theta_b S_{\tau(i)}}.$$

Then

$$(A.12) \qquad \sup_{k \geq 1} E_\nu(e^{\theta_b S_k}\mathbf{1}_{\{k \in A\}}) = \sup_{k \geq 1} Q^*\{\tau(i) = k \text{ for some } i\} \leq 1.$$

From (A.7), (A.10), (A.11) and

$$
\begin{aligned}
E_x(e^{\theta_b S_j}\mathbf{1}_{\{j \in A\}}) &= E_x(e^{\theta_b S_\tau}\mathbf{1}_{\{\tau=j\}}) \\
&\quad + \sum_{h=1}^{j-1} E_x(e^{\theta_b S_\tau}\mathbf{1}_{\{\tau=h\}})E_\nu(e^{\theta_b S_{j-h}}\mathbf{1}_{\{j-h \in A\}}) \\
&\leq E_x(e^{\theta_b S_\tau})\{1 + \sup_{k \geq 1} E_\nu(e^{\theta_b S_k}\mathbf{1}_{\{k \in A\}})\},
\end{aligned}
$$

(3.54) follows from (A.12). □

REFERENCES

[1] BAKER, J.E. (1985). Adaptive selection methods for genetic algorithms. In *Proc. International Conference on Genetic Algorithms and Their Applications* (J. Grefenstette, ed.) 101–111. Erlbaum, Mahwah, NJ.

[2] BAKER, J.E. (1987). Reducing bias and inefficiency in the selection algorithm. In *Genetic Algorithms and Their Applications* (J. Grefenstette, ed.) 14–21. Erlbaum, Mahwah, NJ.

[3] BLANCHET, J. and GLYNN, P. (2008). Efficient rare event simulation for the maximum of heavy-tailed random variables. *Ann. Appl. Probab.* **18** 1351–1378.

[4] BROWN, L. (1986). *Fundamentals of Statistical Exponential Families* **9** IMS Lecture Notes, Hayward, CA.

[5] BUCKLEW, J.A., NEY, P. and SADOWSKY, J.S. (1990). Monte Carlo simulation and large deviation theory for uniformly recurrent Markov chains. *J. Appl. Probab.* **27** 44–59.

[6] CHAN, H.P. and LAI, T.L. (2000). Asymptotic approximations for error probabilities of sequential or fixed sample size tests in exponential families. *Ann. Statist.* **28** 1638–1669.

[7] CHAN, H.P. and LAI, T.L. (2003). Saddlepoint approximations and nonlinear boundary crossing probabilities of Markov random walks. *Ann. Appl. Probab.* **13**, 395–429.

[8] CHAN, H.P. and LAI, T.L. (2007). Efficient importance sampling for Monte Carlo evaluation of exceedance probabilities. *Ann. Appl. Probab.* **17** 440–473.

[9] CHAN, H.P. and LAI, T.L. (2008). A general theory of particle filters in hidden Markov models and some applications. Tech. Report, Department of Statistics, Stanford University.

[10] COLLAMORE, J.F. (2002). Importance sampling techniques for the multidimensional ruin problem for general Markov additive sequences of random vectors. *Ann. Appl. Probab.* **12** 382–421.

[11] CRISAN, D., DEL MORAL, P. and LYONS, T. (1999). Discrete filtering using branching and interacting particle systems. *Markov Process. Related Fields* **5** 293–318.

[12] DE LA PENA, V., LAI, T.L. and SHAO, Q. (2009). *Self-Normalized Processes.* Springer, New York.

[13] DEL MORAL, P., DOUCET, A. and JASRA, A. (2006). Sequential Monte Carlo samplers. *J. Roy. Statist. Soc.* B **68** 411–436.

[14] DEL MORAL, P. and GARNIER, J. (2005). Genealogical particle analysis of rare events. *Ann. Appl. Probab.* **15** 2496–2534.

[15] DEL MORAL, P. and JACOD, J. (2001). Interacting particle filtering with discrete observations. In *Sequential Monte Carlo Methods in Practice* (A. Doucet, N. de Freitas, N. Gordon eds.) 43–75. Springer, New York.

[16] DUPUIS, P. and WANG, H. (2005). Dynamic importance sampling for uniformly recurrent Markov chains. *Ann. Appl. Probab.* **15** 1–38.

[17] DUPUIS, P. and WANG, H. (2007). Subsolutions of an Isaacs equation and and efficient schemes for importance sampling. *Math. Oper. Research* **32** 723–757.

[18] GLASSERMAN, P. and WANG, Y. (1997). Counterexamples in importance sampling for large deviation probabilities. *Ann. Appl. Probab.* **7** 731–746.

[19] NEY, P. and NUMMELIN, E. (1987). Markov additive processes. I. Eigenvalues properties and limit theorems, II. Large deviations. *Ann. Probab.* **15** 561–592, 593–609.

[20] PRESS, N.H., FLANNERY, B.P., TEUKOLSKY, S.A. and VITTERLING, W.T. (1992).
      Numerical Recipes in C: The Art of Scientific Computing, 2nd ed., Cambridge Uni-
      versity Press.

[21] SADOWSKY, J.S. and BUCKLEW, J.A. (1990). On large deviations theory and asymp-
      totically efficient Monte Carlo estimation. *IEEE Trans. Inform. Theory* **36** 579–588.

[22] SIEGMUND, D. (1975). A note on the error probabilities and average sample number
      of the sequential probability ratio test. *J. Roy. Statist. Soc. Ser.* B **37** 394–401.

SEQUENTIAL MONTE CARLO FOR TAIL PROBABILITIES 31

TABLE 1
Monte Carlo estimates of $P\{g(S_n/n) \geq 1/\sqrt{2}\}$.

n	15	20	25
SISR	$(1.10 \pm .07) \times 10^{-3}$	$(1.9 \pm .2) \times 10^{-4}$	$(4.0 \pm .7) \times 10^{-5}$
Direct	$(.9 \pm .3) \times 10^{-3}$	$(1 \pm 1) \times 10^{-4}$	0

TABLE 2
Monte Carlo estimates of $P_0\{S_n/n \geq 2.5\}$.

		n		
	θ	15	20	25
SISR	.1	$(9.68 \pm 1.37) \times 10^{-4}$	$(2.81 \pm .57) \times 10^{-4}$	$(4.70 \pm 1.22) \times 10^{-5}$
	.2	$(9.65 \pm .75) \times 10^{-4}$	$(2.45 \pm .24) \times 10^{-4}$	$(6.70 \pm .64) \times 10^{-5}$
	.273	$(8.31 \pm .48) \times 10^{-4}$	$(2.42 \pm .19) \times 10^{-4}$	$(6.33 \pm .44) \times 10^{-5}$
	.3	$(9.11 \pm .51) \times 10^{-4}$	$(2.54 \pm .20) \times 10^{-4}$	$(5.27 \pm .38) \times 10^{-5}$
	.4	$(9.78 \pm .80) \times 10^{-4}$	$(2.60 \pm .20) \times 10^{-4}$	$(6.58 \pm .67) \times 10^{-5}$
Direct		$(8 \pm 3) \times 10^{-4}$	$(3 \pm 2) \times 10^{-4}$	0

DEPARTMENT OF STATISTICS
AND APPLIED PROBABILITY
6 SCIENCE DRIVE 2
SINGAPORE 117546
EMAIL: stachp@nus.edu.sg

STATISTICS DEPARTMENT
390 SERRA MALL
STANFORD CA 94305
EMAIL:lait@stanford.edu