Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

EFFICIENCY OF ESTIMATING FUNCTIONS OF MEANS IN RARE-EVENT CONTEXTS

Marvin K. Nakayama

Computer Science Department
New Jersey Institute of Technology
Newark, NJ 07102, USA

Bruno Tuffin

Inria, Univ Rennes, CNRS, IRISA
Campus de Beaulieu, 263 Avenue Général Leclerc
35042 Rennes, FRANCE

ABSTRACT

When estimating a function of means, where some but not necessarily all of them correspond to rare events, we provide conditions under which having efficient estimators of each individual mean leads to an efficient estimator of the function of the means. We illustrate this setting through several examples, and numerical results complement the theory.

1 INTRODUCTION

Many rare-event simulation problems consider an estimand expressed in terms of several quantities, where the behaviors of some but not necessarily all of their estimators are critically influenced by the rarity of the event of interest. For example, for a stable GI/GI/1 queue, the expected time for the queue-length process to first hit a large threshold can be written as a ratio. The denominator is the probability (i.e., expectation of an indicator) of hitting the large threshold within a busy period, which is a small rare-event probability that is difficult to estimate. But the numerator, which is the expectation of the minimum of the busy cycle length and the time to hit the large threshold, turns out to be an easily estimated non-rare mean. Thus, the numerator may be efficiently estimated via naive Monte Carlo, but the denominator requires applying a variance-reduction technique (Asmussen and Glynn 2007, Chapters V and VI) to efficiently estimate.

This paper considers estimating a function of means, some (but not necessarily all) pertaining to rare events. We provide conditions that ensure when a good estimator for each mean leads to an efficient estimator of the function of means.

The rest of the paper unfolds as follows. Section 2 reviews the concept of relative error and different notions of efficiency for simulation estimators for problems involving rare events. Section 3 describes the setting of the estimand as a function of means, and presents our main theorem. Throughout these sections, we provide examples to motivate the ideas and notation. Section 4 gives numerical results complementing the theory, and concluding remarks appear in Section 5.

2 RELATIVE ERROR

Let α be a performance measure or estimand for a stochastic model. We consider an estimator $\widehat{\alpha}_n$ of α based on a sample of size n when using some Monte Carlo (MC) method, such as naive MC (simple random sampling) or a variance-reduction technique. Suppose that $\widehat{\alpha}_n$ obeys a *central limit theorem* (CLT)

$$\sqrt{n}\,[\widehat{\alpha}_n - \alpha] \Rightarrow \mathcal{N}(0,\tau^2) \qquad \text{as } n \to \infty, \tag{1}$$

where $\Rightarrow$ denotes convergence in distribution (Billingsley 1995, Section 25), $\mathcal{N}(a,b^2)$ represents a normal random variable with mean a and variance b^2, and τ^2 is the *asymptotic variance* of the CLT (1). If we have a consistent estimator $\widehat{\tau}_n^2$ of τ^2 in the sense that $\widehat{\tau}_n^2 \Rightarrow \tau^2$ as $n \to \infty$, then for fixed large n, we can construct an approximate 95% *confidence interval* (CI) for α based on (1) as $[\widehat{\alpha}_n \pm 1.96\widehat{\tau}_n/\sqrt{n}]$, whose relative half-width is roughly $1.96\tau/[\alpha\sqrt{n}]$. This motivates assessing the quality of the estimator $\widehat{\alpha}_n$ through its *relative error* (RE; e.g., L'Ecuyer et al. 2010), defined for a fixed n and $\alpha \neq 0$ as

$$\mathrm{RE}[\widehat{\alpha}_n] = \frac{\tau}{|\alpha|}. \tag{2}$$

Thus, obtaining a 95% CI with relative half-width w (e.g., $w = 0.1$ for 10% relative half-width) requires a sample size $n \approx (1.96\tau)^2/(w\alpha)^2$, which is $(1.96/w)^2$ times the squared RE.

Example 1 (Estimating an expectation via a sample mean) When $\alpha = \mathbb{E}[Z]$ is a mean of a random variable Z that can be simulated, we can draw n independent and identically distributed (i.i.d.) observations $Z_1, Z_2, \ldots, Z_n$ of Z, and compute the sample average $\widehat{\alpha}_n = (1/n)\sum_{k=1}^{n} Z_k$ as an estimator of α. Then the CLT (1) holds for $\tau^2 = \mathrm{Var}[Z] \in (0,\infty)$ (Billingsley 1995, Theorem 27.1), where $\mathrm{Var}[Z]$ denotes the variance of Z. This set-up also allows for applying importance sampling (IS; Section V.1 of Asmussen and Glynn 2007), as follows. Suppose that $\alpha = \mathbb{E}_0[Q]$ for some random variable Q, where $\mathbb{E}_0[\cdot]$ is the expectation operator under the original distribution F_0 of Q. Now let F_1 be another distribution such that the probability measure corresponding to F_0 is absolutely continuous with respect to the measure for F_1. Then for $\mathbb{E}_1[\cdot]$ as the expectation operator under F_1, we may write $\alpha = \int Q\,\mathrm{d}F_0 = \int Q \frac{\mathrm{d}F_0}{\mathrm{d}F_1}\,\mathrm{d}F_1 = \mathbb{E}_1[QL]$ with $L = \mathrm{d}F_0/\mathrm{d}F_1$ as the likelihood ratio, so $Z = QL$ and $\mathbb{E}[\cdot]$ corresponds to $\mathbb{E}_1[\cdot]$.

When α relates to a rare event, the relative error of an estimator can become quite large as the event of interest becomes rarer. For example, suppose we want to estimate the probability that a real-valued random variable $W_{\langle r\rangle}$ lies in some set $A_{\langle r\rangle} \subset \Re$, where the distribution of $W_{\langle r\rangle}$ and the set $A_{\langle r\rangle}$ may depend on some "rarity parameter" r. In the context of Example 1, we take $Z \equiv Z_{\langle r\rangle} = I(W_{\langle r\rangle} \in A_{\langle r\rangle})$, where $I(\cdot)$ denotes the indicator function. Thus, we consider a sequence of models indexed by r, and examine what happens as $r \to \infty$, where we assume the event $\{W_{\langle r\rangle} \in A_{\langle r\rangle}\}$ becomes "rarer" as r grows in the sense that $\alpha_{\langle r\rangle} \equiv \mathbb{E}[Z_{\langle r\rangle}] = \mathbb{P}(W_{\langle r\rangle} \in A_{\langle r\rangle}) \to 0$ as $r \to \infty$. For example, in a stable GI/GI/1 queue, r might represent a large buffer threshold, and $W_{\langle r\rangle}$ is the maximum queue length before first emptying out; letting $A_{\langle r\rangle} = \{r, r+1, \ldots\}$, we have $\alpha_{\langle r\rangle} = \mathbb{P}(W_{\langle r\rangle} \in A_{\langle r\rangle})$ is the probability that the queue length hits level r during a busy cycle, where $\alpha_{\langle r\rangle} \to 0$ as $r \to \infty$ (Sadowsky 1991, Theorem 1). For each fixed r, the CLT (1) holds for Example 1 with $Z_{\langle r\rangle} = I(W_{\langle r\rangle} \in A_{\langle r\rangle})$ and asymptotic variance $\tau^2 \equiv \tau^2_{\langle r\rangle} = \alpha_{\langle r\rangle}(1-\alpha_{\langle r\rangle})$. Let $\widehat{\alpha}_n \equiv \widehat{\alpha}_{\langle r\rangle,n}$ be the average of a fixed number n of i.i.d. copies of $Z_{\langle r\rangle}$, whose relative error in (2) behaves as

$$\mathrm{RE}[\widehat{\alpha}_{\langle r\rangle,n}] = \frac{\tau_{\langle r\rangle}}{\alpha_{\langle r\rangle}} = \frac{\sqrt{\alpha_{\langle r\rangle}(1-\alpha_{\langle r\rangle})}}{\alpha_{\langle r\rangle}} = \frac{\sqrt{1-\alpha_{\langle r\rangle}}}{\sqrt{\alpha_{\langle r\rangle}}} \to \infty \quad \text{as } r \to \infty.$$

Thus, the relative error of $\widehat{\alpha}_{\langle r\rangle,n}$ blows up as the event of interest becomes rarer, and the required sample size to obtain a fixed relative-width CI grows as roughly $1/\alpha_{\langle r\rangle}$ as $r \to \infty$. This illustrates the primary difficulty with rare-event simulation.

Previous works (e.g., Rubino and Tuffin 2009 and Chapter VI of Asmussen and Glynn 2007) for specific stochastic models devise variance-reduction techniques that produce estimators of such rare-event probabilities $\alpha_{\langle r\rangle}$ for which the relative error grows very slowly, remains bounded or vanishes as $r \to \infty$. Specifically, in Example 1, suppose that for each fixed r, we have an estimator $\widehat{\alpha}_{\langle r\rangle,n}$ of $\alpha_{\langle r\rangle}$ satisfying the CLT (1) with asymptotic variance $\tau^2_{\langle r\rangle}$. Further suppose that $\tau^2_{\langle r\rangle} = O(\alpha^2_{\langle r\rangle})$ as $r \to \infty$, where the notation $b_1(r) = O(b_2(r))$ as $r \to \infty$ for functions b_1 and b_2 means that there exists constants $c_0 > 0$ and $r_0 > 0$ such that $|b_1(r)| \le c_0|b_2(r)|$ for all $r > r_0$. Then the relative error of the estimator $\widehat{\alpha}_{\langle r\rangle,n}$ satisfies

$$\mathrm{RE}[\widehat{\alpha}_{\langle r\rangle,n}] = \frac{\sqrt{O(\alpha^2_{\langle r\rangle})}}{\alpha_{\langle r\rangle}} = \frac{O(\alpha_{\langle r\rangle})}{\alpha_{\langle r\rangle}} = O(1) \quad \text{as } r \to \infty, \tag{3}$$

so the estimator has *bounded relative error* (BRE). Hence, the necessary sample size to obtain a CI of relative half-width w remains bounded as $r \to \infty$. If instead $\tau^2_{\langle r\rangle} = o(\alpha^2_{\langle r\rangle})$ as $r \to \infty$, where the notation $b_1(r) = o(b_2(r))$ means that $\lim_{r\to\infty} b_1(r)/b_2(r) = 0$, then we have

$$\mathrm{RE}[\widehat{\alpha}_{\langle r\rangle,n}] = \frac{\sqrt{o(\alpha^2_{\langle r\rangle})}}{\alpha_{\langle r\rangle}} = \frac{o(\alpha_{\langle r\rangle})}{\alpha_{\langle r\rangle}} = o(1) \quad \text{as } r \to \infty, \tag{4}$$

so the estimator then has *vanishing relative error* (VRE).

We next describe an efficiency property that is slightly weaker than BRE. To do this, define

$$\mathrm{RE}_\zeta[\widehat{\alpha}_{\langle r\rangle,n}] = \frac{\tau_{\langle r\rangle}}{|\alpha_{\langle r\rangle}|^\zeta} \tag{5}$$

for each constant $\zeta > 0$, so (5) with $\zeta = 1$ reduces to just RE in (2). Then when $\alpha_{\langle r\rangle} \to 0$ as $r \to \infty$, we say that the estimator $\widehat{\alpha}_{\langle r\rangle,n}$ of $\alpha_{\langle r\rangle}$ is *logarithmically efficient* (LE) if

$$\limsup_{r\to\infty} \mathrm{RE}_{1-\delta}[\widehat{\alpha}_{\langle r\rangle,n}] = 0 \tag{6}$$

for all $\delta \in (0,1)$. Often a desirable goal when designing an estimator of a small probability based on large-deviations theory, the property (6) is equivalent to $\liminf_{r\to\infty} \frac{|\ln \tau_{\langle r\rangle}|}{|\ln(\alpha_{\langle r\rangle})|} \geq 1$, which shows where LE gets its name. While the literature on rare-event simulation has primarily focused on LE estimators for small probabilities, some estimands of interest have $\alpha_{\langle r\rangle} \to \infty$ as $r \to \infty$, e.g., a stochastic process's expected hitting time to a rarely visited set of states. When instead considering an estimand for which $\alpha_{\langle r\rangle} \to \pm\infty$ as $r \to \infty$, we redefine LE to mean that for all $\delta > 0$,

$$\limsup_{r\to\infty} \mathrm{RE}_{1+\delta}[\widehat{\alpha}_{\langle r\rangle,n}] = 0, \tag{7}$$

which is equivalent to $\limsup_{r\to\infty} \frac{|\ln \tau_{\langle r\rangle}|}{|\ln(\alpha_{\langle r\rangle})|} \leq 1$. We do not define LE for the case when $\lim_{r\to\infty} \alpha_{\langle r\rangle} = \chi_0 \in (-\infty,\infty)$ with $\chi_0 \neq 0$. Indeed, the idea of LE originates from studies of rare-event probabilities through large-deviations theory, where a goal in simulation is to construct an estimator whose standard deviation and mean both converge to 0 at the same exponential rate as $r \to \infty$. While (7) extends this conceptual framework to handle $\alpha_{\langle r\rangle} \to \pm\infty$, an appropriate definition for LE is not obvious when $\alpha_{\langle r\rangle}$ converges to a finite $\chi_0 \neq 0$, so we do not consider this case. We could also take into account the CPU time to compute $\widehat{\alpha}_{\langle r\rangle,n}$ via the work-normalized RE (WNRE), as in L'Ecuyer et al. (2010), but we focus here on only RE.

In what follows, we will sometimes omit the subscript $\langle r\rangle$ on variables to simplify the notation.

3 FUNCTIONS OF MEANS

We will consider estimands α having a specific form:

$$\alpha = g(\boldsymbol{\theta}) \tag{8}$$

for some known function $g : \Re^d \to \Re$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d) \in \Re^d$ is a vector of unknown parameters, for some fixed $d \geq 1$. Both g and d do not depend on the rarity parameter r, but the components θ_i of $\boldsymbol{\theta}$ may. We assume that there is a d-dimensional random vector $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ such that its mean

$$\mathbb{E}[\mathbf{X}] = \boldsymbol{\theta} \tag{9}$$

and that we can generate i.i.d. copies of $\mathbf{X}$.

The random vector $\mathbf{X}$ has some joint distribution. The d components of $\mathbf{X}$ may be dependent, as is often the case when they are generated from the same simulation, or they may be independent, generated from different simulations applying different MC methods. For example, X_1 could be generated using naive MC, and $X_2 = QL$ could be generated (independently of X_1) using IS, as in Example 1. In this case, the expectation operator $\mathbb{E}[\cdot]$ in (9) applies $\mathbb{E}_0[\cdot]$ (resp., $\mathbb{E}_1[\cdot]$) to the first (resp., second) coordinate of $\mathbf{X}$. Goyal et al. (1992) call this approach *measure-specific IS* (MSIS).

Many estimands of interest have the form in (8).

Example 2 (Expected hitting time) Consider a nondelayed regenerative process $Y = (Y(t) : t \geq 0)$ evolving on a state space $S \subset \Re^v$ for some $v \geq 1$ (Kalashnikov 1994), so the process "probabilistically restarts" at an infinite sequence of regeneration times $0 = T_{-1} < T_0 < T_1 < \cdots$. We can split the sample path of Y into i.i.d. *regenerative cycles* demarcated by successive pairs of regeneration times. For example, if Y represents the queue-length process (including any customer undergoing service) in a stable GI/GI/1 queue, then regenerations occur when a customer arrives to an empty system, where the initial customer arrives at time $t = 0$ to an empty system. If Y is an irreducible and positive-recurrent continuous-time Markov chain (CTMC), returns to any fixed state $s_0 \in S$ constitute regenerations. Also consider a set $A \subset S$ (with $s_0 \notin A$ in the case of a CTMC), and let $T_A = \inf\{t \geq 0 : Y(t) \in A\}$ be the hitting time to A. For example, in a GI/GI/1 queue, letting $A \equiv A_{\langle r \rangle} = \{r, r+1, \ldots\}$ for some (large) $r > 0$ results in T_A denoting the first time the queue length hits r. Instead, if Y represents a Markovian reliability system, which will be covered in Example 5, A could be the set of "system failed" states, making T_A the first time to system failure. We are interested in estimating $\alpha = \mathbb{E}[T_A]$. As shown in, e.g., Goyal et al. (1992) and Glynn et al. (2017), we can express α as a ratio $\alpha = \eta/\gamma$, with $\eta = \mathbb{E}[\min(T_A, T_0)]$ and $\gamma = \mathbb{E}[I(T_A < T_0)] = \mathbb{P}(T_A < T_0)$, where both $\min(T_A, T_0)$ and $I(T_A < T_0)$ are measurable with respect to the sigma-field of the process Y up to time T_0. Then for the random vector $\mathbf{X} = (X_1, X_2)$ in (9) with dimension $d = 2$, where $X_1 = \min(T_A, T_0)$ and $X_2 = I(T_A < T_0)$, (8) has $\boldsymbol{\theta} = (\eta, \gamma)$, and $g(x_1, x_2) = x_1/x_2$. When Y is a CTMC, we can obtain reduced-variance estimators of η, γ, and α via *conditional MC* (CMC; e.g., Section V.4 of Asmussen and Glynn 2007) by conditioning on the CTMC's embedded discrete-time Markov chain (DTMC). A simulation applying CMC then generates only the embedded DTMC, with the exponential holding times in visited states replaced by their conditional means. (Later examples involving CTMCs can also apply CMC in this manner.)

Example 3 (Derivative of expected hitting time) Suppose for the CTMC Y in Example 2, its infinitesimal generator matrix is parameterized by λ, which may denote the transition rate of a specific group of transitions. For example, λ may represent the failure rate of a component in a reliability setting, or the arrival rate of customers to a queue. In this case, we write $\beta \equiv \mathbb{E}[T_A]$ as $\beta(\lambda) = \eta(\lambda)/\gamma(\lambda)$ as these values now depend on λ. We are interested in computing the derivative $\beta'(\lambda) \equiv \frac{\mathrm{d}}{\mathrm{d}\lambda}\beta(\lambda)$, which satisfies $\beta'(\lambda) = [\eta'(\lambda)\gamma(\lambda) - \eta(\lambda)\gamma'(\lambda)]/\gamma^2(\lambda)$, where $\eta'(\lambda) = \frac{\mathrm{d}}{\mathrm{d}\lambda}\eta(\lambda)$ and $\gamma'(\lambda) = \frac{\mathrm{d}}{\mathrm{d}\lambda}\gamma(\lambda)$. Using the likelihood ratio derivative method (Glynn 1990), we can express $\eta'(\lambda) = \mathbb{E}[\min(T_A, T_0)L']$ and $\eta'(\lambda) = \mathbb{E}[I(T_A < T_0)L']$, where L' is the derivative with respect to λ of the likelihood ratio L up to $\min(T_A, T_0)$; see Nakayama (1998) for details when further applying CMC. Thus, for random vector $\mathbf{X} = (X_1, X_2, X_3, X_4)$ in (9) with dimension $d = 4$, and $X_1 = \min(T_A, T_0)$ $X_2 = I(T_A < T_0)$, $X_3 = \min(T_A, T_0)L'$, and $X_4 = I(T_A < T_0)L'$, we can write $\beta'(\lambda)$ as α in (8) using $\boldsymbol{\theta} = (\eta(\lambda), \gamma(\lambda), \eta'(\lambda), \gamma'(\lambda))$, and $g(x_1, x_2, x_3, x_4) = [x_3 x_2 - x_4 x_1]/x_2^2$.

Example 4 (Conditional expectation) Let (Z_1, Z_2) be a random vector, and let $A \subset \Re$ be a set. Then the conditional expectation $\alpha = \mathbb{E}[Z_1 \mid Z_2 \in A] = \mathbb{E}[Z_1 I(Z_2 \in A)]/\mathbb{E}[I(Z_2 \in A)]$ may be expressed in the form in (8) by considering the random vector $\mathbf{X} = (X_1, X_2)$ in (9) with dimension $d = 2$, $X_1 = Z_1 I(Z_2 \in A)$, $X_2 = I(Z_2 \in A)$, $\boldsymbol{\theta} = (\mathbb{E}[X_1], \mathbb{E}[X_2])$, and $g(x_1, x_2) = x_1/x_2$.

As in Section 2, we consider a sequence of problems indexed by rarity parameter r, and take the limit as $r \to \infty$. In (8), we assume that the value of the unknown parameter $\boldsymbol{\theta} \equiv \boldsymbol{\theta}_{\langle r \rangle} = (\theta_{\langle r \rangle,1}, \theta_{\langle r \rangle,2}, \ldots, \theta_{\langle r \rangle,d})$ may depend on r, but the function g does not. Then the estimand $\alpha \equiv \alpha_r$ in (8) also depends on r, and we want to study when $\alpha_{\langle r \rangle}$ can be estimated with BRE as in (3), with VRE as in (4), or with LE as in (6) or (7). As noted before in Section 2, we will sometimes omit the subscript $\langle r \rangle$ to simplify notation.

To estimate $\alpha_{\langle r \rangle} = g(\boldsymbol{\theta}_{\langle r \rangle})$ in (8), we generate $n \geq 1$ i.i.d. observations $\mathbf{X}_{\langle r \rangle,k}$ $k = 1,2,\ldots,n$, of $\mathbf{X} \equiv \mathbf{X}_{\langle r \rangle} = (X_{\langle r \rangle,1}, X_{\langle r \rangle,2}, \ldots, X_{\langle r \rangle,d})$, whose joint distribution may depend on r. By (9), we then obtain an unbiased estimator $\widehat{\boldsymbol{\theta}}_{\langle r \rangle,n} = (\widehat{\theta}_{\langle r \rangle,n,1}, \widehat{\theta}_{\langle r \rangle,n,2}, \ldots, \widehat{\theta}_{\langle r \rangle,n,d}) = (1/n)\sum_{k=1}^{n} \mathbf{X}_{\langle r \rangle,k}$ of $\boldsymbol{\theta} = \boldsymbol{\theta}_{\langle r \rangle} = \mathbb{E}[\mathbf{X}_{\langle r \rangle}]$, yielding

$$\widehat{\alpha}_{\langle r \rangle,n} = g(\widehat{\boldsymbol{\theta}}_{\langle r \rangle,n}) \tag{10}$$

as a plug-in estimator of $\alpha_{\langle r \rangle}$. As noted before, this setup allows for applying MSIS. For instance, Example 2 can take $X_{\langle r \rangle,1} = \min(T_A, T_0)$ generated via naive MC so $\theta_{\langle r \rangle,1} = \mathbb{E}_0[\min(T_A, T_0)]$, and $X_{\langle r \rangle,2} = Q_{\langle r \rangle} L_{\langle r \rangle}$ sampled using IS with $Q_{\langle r \rangle} = I(T_{A_{\langle r \rangle}} < T_{0,\langle r \rangle})$ and $L_{\langle r \rangle}$ the likelihood ratio, so $\theta_{\langle r \rangle,2} = \mathbb{E}_0[Q_{\langle r \rangle}] = \mathbb{E}_1[Q_{\langle r \rangle} L_{\langle r \rangle}]$.

For the estimator $\widehat{\alpha}_{\langle r \rangle,n}$ in (10), we will study its relative error as in (2). We want to determine conditions under which the RE properties of an estimator $\widehat{\boldsymbol{\theta}}_{\langle r \rangle,n}$ of $\boldsymbol{\theta}_{\langle r \rangle}$ will carry over to the plug-in estimator $\widehat{\alpha}_{\langle r \rangle,n} = g(\widehat{\boldsymbol{\theta}}_{\langle r \rangle,n})$ of $\alpha_{\langle r \rangle}$. Let $\Sigma_{\langle r \rangle} = (\Sigma_{\langle r \rangle,i,j} : i,j = 1,2,\ldots,d)$ be the covariance matrix of the random vector $\mathbf{X}_{\langle r \rangle}$, where $\Sigma_{\langle r \rangle,i,j} = \mathrm{Cov}[X_{\langle r \rangle,i}, X_{\langle r \rangle,j}] = \mathbb{E}[(X_{\langle r \rangle,i} - \mathbb{E}[X_{\langle r \rangle,i}])(X_{\langle r \rangle,j} - \mathbb{E}[X_{\langle r \rangle,j}])]$. Also, let $\nabla g(\boldsymbol{\theta}) = (g_1(\boldsymbol{\theta}), g_2(\boldsymbol{\theta}), \ldots, g_d(\boldsymbol{\theta}))$ be the gradient (when it exists) of $g(\boldsymbol{\theta})$, where $g_i(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_i} g(\boldsymbol{\theta})$, $i = 1,2,\ldots,d$. For each $\psi, \zeta > 0$, define

$$h_{i,\psi,\zeta}(\boldsymbol{\theta}_{\langle r \rangle}) = \left| \frac{\theta_{\langle r \rangle,i}^{\psi} g_i(\boldsymbol{\theta}_{\langle r \rangle})}{g^{\zeta}(\boldsymbol{\theta}_{\langle r \rangle})} \right|. \tag{11}$$

For analyzing LE in (6) and (7) (but unneeded for handling BRE or VRE), we further impose the following.

Assumption A1 As $r \to \infty$, $\alpha_{\langle r \rangle} \to \chi_0 \in \{0, \infty, -\infty\}$ and $\theta_{\langle r \rangle,i} \to \chi_i \in [-\infty, \infty]$ for each $i = 1,2,\ldots,d$.

Assumption A1 disallows the estimand $\alpha_{\langle r \rangle}$ to converge to a nonzero finite constant when establishing LE. This is to retain the spirit of the framework that LE was originally defined to consider, as discussed after (7). But for $i = 1,2,\ldots,d$, we permit some of the $\theta_{\langle r \rangle,i}$ to converge to nonzero finite constants χ_i. This allows us to handle a situation that occurs in Example 2 when estimating the expected time for the queue-length process of a stable GI/GI/1 queue to first hit a high level r; this will be explained later in Example 8. For each $\delta \in (0,1)$ and $i = 0,1,\ldots,d$, define δ_i based on Assumption A1 as follows:

$$\delta_i = \begin{cases} -\delta & \text{if } \chi_i = 0, \\ 0 & \text{if } \chi_i \in (-\infty, \infty) \text{ with } \chi_i \neq 0, \\ \delta & \text{if } |\chi_i| = \infty. \end{cases} \tag{12}$$

Also, define $\mathbf{0} = (0,0,\ldots,0) \in \Re^d$, the vector of all 0s.

Theorem 1 Assume that for each r, the function g in (8) is differentiable at $\boldsymbol{\theta}_{\langle r \rangle}$ with $\nabla g(\boldsymbol{\theta}_{\langle r \rangle}) \neq \mathbf{0}$ and $\Sigma_{\langle r \rangle}$ is finite and positive definite. Then the following hold as $r \to \infty$:

(i) If $h_{i,1,1}(\boldsymbol{\theta}_{\langle r \rangle})\mathrm{RE}[\widehat{\theta}_{\langle r \rangle,n,i}] = O(1)$ as $r \to \infty$ for all $i = 1,2,\ldots,d$, then $\widehat{\alpha}_{\langle r \rangle,n}$ has BRE.

(ii) If $h_{i,1,1}(\boldsymbol{\theta}_{\langle r \rangle})\mathrm{RE}[\widehat{\theta}_{\langle r \rangle,n,i}] = o(1)$ as $r \to \infty$ for all $i = 1,2,\ldots,d$, then $\widehat{\alpha}_{\langle r \rangle,n}$ has VRE.

(iii) If Assumption A1 holds, and for each $i = 1,2,\ldots,d$, if

$$\limsup_{r \to \infty} h_{i,1+\delta_i,1+\delta_0}(\boldsymbol{\theta}_{\langle r \rangle})\mathrm{RE}_{1+\delta_i}[\widehat{\theta}_{\langle r \rangle,n,i}] = 0 \tag{13}$$

for all $\delta \in (0,1)$ with δ_i in (12), then $\widehat{\alpha}_{\langle r \rangle,n}$ is LE.

In Theorem 1(i) and (ii), if $h_{i,1,1}(\alpha_{\langle r \rangle})$ is bounded as $r \to \infty$ for all $i = 1,2,\ldots,d$, then the overall estimator $\widehat{\alpha}_{\langle r \rangle,n}$ has BRE or VRE when each $\widehat{\theta}_{\langle r \rangle,n,i}$ has that same property. To understand Theorem 1(iii) for LE, consider the special case when $h_{i,1+\delta_i,1+\delta_0}(\alpha_{\langle r \rangle})$ is bounded as $r \to \infty$ for all $i = 1,2,\ldots,d$. Note that if $\chi_i \in \{0, -\infty, \infty\}$, then $\limsup_{r\to\infty} \mathrm{RE}_{1+\delta_i}[\widehat{\theta}_{\langle r\rangle,n,i}] = 0$ means that $\widehat{\theta}_{\langle r\rangle,n,i}$ is LE, as in (6) or (7); if instead $\chi_i \in (-\infty,\infty)$ with $\chi_i \neq 0$, then because (12) implies $\delta_i = 0$ in (13), $\limsup_{r\to\infty} \mathrm{RE}_{1+\delta_i}[\widehat{\theta}_{\langle r\rangle,n,i}] = \limsup_{r\to\infty} \mathrm{RE}[\widehat{\theta}_{\langle r\rangle,n,i}] = 0$ means that $\widehat{\theta}_{\langle r\rangle,n,i}$ has VRE. Thus, if $h_{i,1+\delta_i,1+\delta_0}(\alpha_{\langle r\rangle})$ is bounded as $r\to\infty$ for all $i = 1,2,\ldots,d$, and Assumption A1 holds, then the overall estimator $\widehat{\alpha}_{\langle r\rangle,n}$ is LE when each $\widehat{\theta}_{\langle r\rangle,n,i}$ is LE. But when $\chi_i \in (-\infty,\infty)$ with $\chi_i \neq 0$, Theorem 1 also allows for $\widehat{\theta}_{\langle r\rangle,n,i}$ to be just BRE and not VRE if $h_{i,1+\delta_i,1+\delta_0}(\alpha_{\langle r\rangle}) \to 0$ as $r\to\infty$, which we will later see occurs in Example 8.

Proof of Theorem 1. For each fixed r, the condition that the covariance matrix $\Sigma_{\langle r\rangle}$ is finite and positive definite ensures that $\widehat{\boldsymbol{\theta}}_{\langle r\rangle,n}$ obeys a multivariate CLT (Billingsley 1995, Theorem 29.5)

$$\sqrt{n}\left[\widehat{\boldsymbol{\theta}}_{\langle r\rangle,n} - \boldsymbol{\theta}_{\langle r\rangle}\right] \Rightarrow \mathscr{N}_d(\mathbf{0},\Sigma_{\langle r\rangle}) \quad \text{as } n\to\infty, \tag{14}$$

where $\mathscr{N}_d(\mathbf{0},\Sigma_{\langle r\rangle})$ is a d-dimensional normal random vector with mean vector $\mathbf{0}$ and covariance matrix $\Sigma_{\langle r\rangle}$. Under our assumption that $\nabla g(\boldsymbol{\theta}_{\langle r\rangle}) \neq \mathbf{0}$, the delta method (Serfling 1980, p. 124) then yields the CLT

$$\sqrt{n}\left[\widehat{\alpha}_{\langle r\rangle,n} - \alpha_{\langle r\rangle}\right] \Rightarrow \mathscr{N}(0,\tau^2_{\langle r\rangle}) \quad \text{as } n\to\infty, \tag{15}$$

$$\text{with } \tau^2_{\langle r\rangle} = \sum_{i=1}^d\sum_{j=1}^d g_i(\boldsymbol{\theta}_{\langle r\rangle}) g_j(\boldsymbol{\theta}_{\langle r\rangle})\Sigma_{\langle r\rangle,i,j} \tag{16}$$

as the asymptotic variance. The Cauchy-Schwarz inequality implies that $|\Sigma_{\langle r\rangle,i,j}| \leq \sqrt{\mathrm{Var}[X_{\langle r\rangle,i}]\mathrm{Var}[X_{\langle r\rangle,j}]}$, and also we have that $\mathrm{RE}[\widehat{\theta}_{\langle r\rangle,n,i}] = \sqrt{\mathrm{Var}[X_{\langle r\rangle,i}]}/|\theta_{\langle r\rangle,i}|$, as in (2). Thus, the RE of $\widehat{\alpha}_{\langle r\rangle,n}$ then satisfies

$$\begin{aligned}
\mathrm{RE}[\widehat{\alpha}_{\langle r\rangle,n}] &= \left(\sum_{i=1}^d\sum_{j=1}^d \frac{g_i(\boldsymbol{\theta}_{\langle r\rangle})g_j(\boldsymbol{\theta}_{\langle r\rangle})}{g^2(\boldsymbol{\theta}_{\langle r\rangle})}\Sigma_{\langle r\rangle,i,j}\right)^{1/2} \leq \left(\sum_{i=1}^d\sum_{j=1}^d \left|\frac{g_i(\boldsymbol{\theta}_{\langle r\rangle})g_j(\boldsymbol{\theta}_{\langle r\rangle})}{g^2(\boldsymbol{\theta}_{\langle r\rangle})}\Sigma_{\langle r\rangle,i,j}\right|\right)^{1/2} \\
&\leq \left(\sum_{i=1}^d\sum_{j=1}^d \left|\frac{g_i(\boldsymbol{\theta}_{\langle r\rangle})g_j(\boldsymbol{\theta}_{\langle r\rangle})}{g^2(\boldsymbol{\theta}_{\langle r\rangle})}\right|\sqrt{\mathrm{Var}[X_{\langle r\rangle,i}]\mathrm{Var}[X_{\langle r\rangle,j}]}\right)^{1/2} \\
&= \left(\sum_{i=1}^d\sum_{j=1}^d \left|\frac{\theta_{\langle r\rangle,i}g_i(\boldsymbol{\theta}_{\langle r\rangle})\theta_{\langle r\rangle,j}g_j(\boldsymbol{\theta}_{\langle r\rangle})}{g^2(\boldsymbol{\theta}_{\langle r\rangle})}\right|\mathrm{RE}[\widehat{\theta}_{\langle r\rangle,n,i}]\mathrm{RE}[\widehat{\theta}_{\langle r\rangle,n,j}]\right)^{1/2} \\
&= \left(\sum_{i=1}^d\sum_{j=1}^d h_{i,1,1}(\boldsymbol{\theta}_{\langle r\rangle})\,\mathrm{RE}[\widehat{\theta}_{\langle r\rangle,n,i}]\,h_{j,1,1}(\boldsymbol{\theta}_{\langle r\rangle})\,\mathrm{RE}[\widehat{\theta}_{\langle r\rangle,n,j}]\right)^{1/2}
\end{aligned} \tag{17}$$

by (11), so parts (i) and (ii) hold by our assumptions about each $h_{i,1,1}(\boldsymbol{\theta}_{\langle r\rangle})\mathrm{RE}[\widehat{\theta}_{\langle r\rangle,n,i}]$ being $O(1)$ or $o(1)$.

For part (iii) under Assumption A1, we simplify the discussion by considering the square of (5) for $\zeta = 1+\delta_0$. As in (17), we get

$$\begin{aligned}
\left(\mathrm{RE}_{1+\delta_0}[\widehat{\alpha}_{\langle r\rangle,n}]\right)^2 = \frac{\tau^2_{\langle r\rangle}}{|\alpha_{\langle r\rangle}|^{2+2\delta_0}} &\leq \sum_{i=1}^d\sum_{j=1}^d \left|\frac{g_i(\boldsymbol{\theta}_{\langle r\rangle})g_j(\boldsymbol{\theta}_{\langle r\rangle})}{g^{2+2\delta_0}(\boldsymbol{\theta}_{\langle r\rangle})}\right|\sqrt{\mathrm{Var}[X_{\langle r\rangle,i}]\mathrm{Var}[X_{\langle r\rangle,j}]} \\
&= \sum_{i=1}^d\sum_{j=1}^d \left|\frac{\theta^{1+\delta_i}_{\langle r\rangle,i}g_i(\boldsymbol{\theta}_{\langle r\rangle})\theta^{1+\delta_j}_{\langle r\rangle,j}g_j(\boldsymbol{\theta}_{\langle r\rangle})}{g^{1+\delta_0}(\boldsymbol{\theta}_{\langle r\rangle})g^{1+\delta_0}(\boldsymbol{\theta}_{\langle r\rangle})}\right|\mathrm{RE}_{1+\delta_i}[\widehat{\theta}_{\langle r\rangle,n,i}]\mathrm{RE}_{1+\delta_j}[\widehat{\theta}_{\langle r\rangle,n,j}] \\
&= \sum_{i=1}^d\sum_{j=1}^d h_{i,1+\delta_i,1+\delta_0}(\boldsymbol{\theta}_{\langle r\rangle})\,\mathrm{RE}_{1+\delta_i}[\widehat{\theta}_{\langle r\rangle,n,i}]\,h_{j,1+\delta_j,1+\delta_0}(\boldsymbol{\theta}_{\langle r\rangle})\,\mathrm{RE}_{1+\delta_j}[\widehat{\theta}_{\langle r\rangle,n,j}]
\end{aligned}$$

by (11), so part (iii) holds by (6) and (7) under our condition (13). □

Even though $\widehat{\boldsymbol{\theta}}_{\langle r\rangle,n}$ is an unbiased estimator of $\boldsymbol{\theta}_{\langle r\rangle}$, we may have that $\widehat{\alpha}_{\langle r\rangle,n} = g(\widehat{\boldsymbol{\theta}}_{\langle r\rangle,n})$ is a *biased* estimator of $\alpha_{\langle r\rangle} = g(\boldsymbol{\theta}_{\langle r\rangle})$ when g is nonlinear, as in Examples 2–4, as well as the ones below. However, for fixed r, we still obtain the CLT in (15) as $n \to \infty$ with the true estimand $\alpha_{\langle r\rangle}$ as the centering constant. For fixed r, the bias contributes negligibly to the mean-square error (MSE) of $\widehat{\alpha}_{\langle r\rangle,n}$ as $n \to \infty$, where the MSE can be decomposed as the sum of the variance and squared bias. Indeed, under appropriate technical conditions (e.g., Section 2.4.1 of Shao and Tu 1995) on g in (8) and the moments of $\mathbf{X}$ in (9), as $n \to \infty$,

$$\mathbb{E}[\widehat{\alpha}_{\langle r\rangle,n}] = \mathbb{E}[g(\widehat{\boldsymbol{\theta}}_{\langle r\rangle,n})] = g(\boldsymbol{\theta}_{\langle r\rangle}) + \frac{1}{n}\sum_{i=1}^{d}\sum_{j=1}^{d} g_i(\boldsymbol{\theta}_{\langle r\rangle}) g_j(\boldsymbol{\theta}_{\langle r\rangle}) \Sigma_{\langle r\rangle,i,j} + o\left(\frac{1}{n}\right),$$

so the squared bias shrinks as $O(1/n^2)$, faster than the estimator's variance, which typically decreases as $\tau^2_{\langle r\rangle}/n + o(1/n)$.

3.1 Examples

We next provide examples (some expanding on our previous ones) that fit into the framework of Theorem 1.

Example 5 (Mean time to failure (MTTF)) We now specialize Example 2 for a highly reliable Markovian system (HRMS), as in Goyal et al. (1992) and Shahabuddin (1994). An HRMS comprises a fixed collection of dependable components subject to random failures and repairs, with component failure and repair times exponentially distributed. The stochastic process Y models the evolution of the HRMS over time, and let S be its state space, where each state $s \in S$ keeps track of the components that are currently up and down, along with any other information (e.g., queueing of failed components at the repair stations) necessary to ensure that Y is a CTMC. The system begins at time 0 in $s_0 \in S$, the state with all components operational. The system fails when certain combinations of components are failed, and let $A \subset S$ denote the set of failed states, where we assume that $s_0 \notin A$, and the time to failure is T_A. To incorporate the rarity parameter r, we model the component failure rates as positive powers of $\varepsilon = 1/r$, and repair rates are constants, and we analyze the HRMS as $r \to \infty$. Hence, as r increases, individual components become more reliable, but the set A of failed states does not change. The MTTF $\alpha_{\langle r\rangle} = \mathbb{E}_0[T_A] \equiv \mathbb{E}_0[T_{A,\langle r\rangle}]$ can be expressed as a ratio $\eta_{\langle r\rangle}/\gamma_{\langle r\rangle}$ with $\eta_{\langle r\rangle} = \mathbb{E}_0[\min(T_{A,\langle r\rangle}, T_{0,\langle r\rangle})]$ and $\gamma_{\langle r\rangle} = \mathbb{E}_0[I(T_{A,\langle r\rangle} < T_{0,\langle r\rangle})] = \mathbb{E}_1[I(T_{A,\langle r\rangle} < T_{0,\langle r\rangle})L]$, where $\mathbb{E}_0$ is the expectation operator under the original system dynamics, and $\mathbb{E}_1[\cdot]$ denotes expectation under IS, with L as the resulting likelihood ratio. We apply MSIS to estimate the ratio, with naive MC for estimating the numerator and IS for the denominator. Therefore, in (9), the random vector $\mathbf{X}$ has dimension $d = 2$, and $\mathbf{X} = (X_1, X_2) = (\min(T_{A,\langle r\rangle}, T_{0,\langle r\rangle}), I(T_{A,\langle r\rangle} < T_{0,\langle r\rangle})L)$, where the expectation operator $\mathbb{E}[\cdot]$ uses $\mathbb{E}_0[\cdot]$ (resp., $\mathbb{E}_1[\cdot]$) on the first (resp., second) coordinate of $\mathbf{X}$. Also, (8)–(10) have $\boldsymbol{\theta}_{\langle r\rangle} = (\eta_{\langle r\rangle}, \gamma_{\langle r\rangle})$, and $g(x_1, x_1) = x_1/x_2$. Since (11) has $h_{1,1,1}(\boldsymbol{\theta}_{\langle r\rangle}) = h_{2,1,1}(\boldsymbol{\theta}_{\langle r\rangle}) = 1$, BRE (resp., VRE) for the MTTF ratio estimator follows if the estimators for $\eta_{\langle r\rangle}$ and $\gamma_{\langle r\rangle}$ both have BRE (resp., VRE) by Theorem 1(i) and (ii). When applying balanced failure biasing (described further in Example 10) for the IS for estimating the denominator $\gamma_{\langle r\rangle}$, Shahabuddin (1994) establishes the BRE of the MTTF MSIS ratio estimator when further applying CMC by conditioning on the embedded DTMC; Nakayama (1996), Rubino and Tuffin (2009) and L'Ecuyer and Tuffin (2012) provide related results for other forms of IS to estimate $\gamma_{\langle r\rangle}$.

Example 6 (Alternative MTTF estimator) Instead of using the ratio estimator of the MTTF in Example 5, Nakayama and Tuffin (2019) propose a different approach, which can be more efficient in certain situations and also applies more generally, as for Example 2. Specifically, they express the MTTF as

$$\begin{aligned}\alpha_{\langle r\rangle} &= \frac{\mathbb{E}_0[\min(T_{A,\langle r\rangle}, T_{0,\langle r\rangle}) I(T_{A,\langle r\rangle} > T_{0,\langle r\rangle})] + \mathbb{E}_0[\min(T_{A,\langle r\rangle}, T_{0,\langle r\rangle}) I(T_{A,\langle r\rangle} < T_{0,\langle r\rangle})]}{\mathbb{E}_0[I(T_{A,\langle r\rangle} < T_{0,\langle r\rangle})]} \\ &= \frac{\mathbb{E}_0[T_{0,\langle r\rangle} I(T_{A,\langle r\rangle} > T_{0,\langle r\rangle})] + \mathbb{E}_1[T_{A,\langle r\rangle} I(T_{A,\langle r\rangle} < T_{0,\langle r\rangle})L]}{\mathbb{E}_1[I(T_{A,\langle r\rangle} < T_{0,\langle r\rangle})L]}.\end{aligned}$$

To put this in the framework of (8)–(10), we apply MSIS with the random vector $\mathbf{X}_{\langle r\rangle} = (X_{\langle r\rangle,1}, X_{\langle r\rangle,2}, X_{\langle r\rangle,3})$ of dimension $d = 3$ and $X_{\langle r\rangle,1} = T_{0,\langle r\rangle} I(T_{A,\langle r\rangle} > T_{0,\langle r\rangle})$, $X_{\langle r\rangle,2} = T_{A,\langle r\rangle} I(T_{A,\langle r\rangle} < T_{0,\langle r\rangle})L$, and $X_{\langle r\rangle,3} = I(T_{A,\langle r\rangle} < T_{0,\langle r\rangle})L$, where $X_{\langle r\rangle,1}$ is generated with naive MC, and $(X_{\langle r\rangle,2}, X_{\langle r\rangle,3})$ is generated with IS. Hence, the expectation operator $\mathbb{E}[\cdot]$ in (9) applies $\mathbb{E}_0[\cdot]$ (resp., $\mathbb{E}_1[\cdot]$) for the first (resp., second and third) coordinate(s), so $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ with $\theta_1 = \mathbb{E}_0[T_{0,\langle r\rangle} I(T_{A,\langle r\rangle} > T_{0,\langle r\rangle})]$, $\theta_2 = \mathbb{E}_1[T_{A,\langle r\rangle} I(T_{A,\langle r\rangle} < T_{0,\langle r\rangle})L]$, and $\theta_3 = \mathbb{E}_1[I(T_{A,\langle r\rangle} < T_{0,\langle r\rangle})L]$, and $g(x_1, x_2, x_3) = (x_1 + x_2)/x_3$ in (8). Since (11) has $h_{1,1,1}(\boldsymbol{\theta}_{\langle r\rangle}) = \frac{\theta_{\langle r\rangle,1}}{\theta_{\langle r\rangle,1} + \theta_{\langle r\rangle,2}} \le 1$, $h_{2,1,1}(\boldsymbol{\theta}_{\langle r\rangle}) = \frac{\theta_{\langle r\rangle,2}}{\theta_{\langle r\rangle,1} + \theta_{\langle r\rangle,2}} \le 1$ and $h_{3,1,1}(\boldsymbol{\theta}_{\langle r\rangle}) = 1$, they are all bounded, so by Theorem 1(i), the estimator of $\alpha_{\langle r\rangle}$ has BRE if the estimators of each of the three components have the same property. We can also establish if the estimator of $\alpha_{\langle r\rangle}$ is LE (which is more appropriate in, e.g., the GI/GI/1 setting) through an analysis that will be used in Example 8.

Example 7 For Example 3 specialized to an HRMS, Nakayama (1998) establishes the BRE of the derivative estimator of the MTTF with respect to a component failure rate when using MSIS with balanced failure biasing under certain assumptions when further applying CMC by conditioning on the embedded DTMC.

Example 8 (Expected time for GI/GI/1 queue length to reach a large threshold) For the queue-length process of a stable GI/GI/1 queue, as briefly discussed in Example 2, let $\mathbb{E}_0$ (resp., $\mathbb{P}_0$) denote the expectation operator (resp., probability measure) under the original system dynamics. We consider the estimand $\alpha_{\langle r\rangle} = \mathbb{E}_0[T_{A_{\langle r\rangle}}]$, which is the expected hitting time to the set $A_{\langle r\rangle} = \{r, r+1, r+2, \ldots\}$. Let U (resp., V) denote a generic interarrival (resp., service) time, with $U, V \ge 0$ independent and assumed to have light-tailed distributions (i.e., their moment-generating functions exist in a neighborhood of the origin, so all moments of U and V are finite) not depending on r. Further assume that $0 < \mathbb{E}_0[V] < \mathbb{E}_0[U]$ and also that $\mathbb{P}_0(V > U) > 0$. Then we have $\alpha_{\langle r\rangle} = \eta_{\langle r\rangle}/\gamma_{\langle r\rangle} \in (0, \infty)$, with $\eta_{\langle r\rangle} = \mathbb{E}_0[\min(T_{A_{\langle r\rangle}}, T_0)]$ and $\gamma_{\langle r\rangle} = \mathbb{E}_0[I(T_{A_{\langle r\rangle}} < T_0)] = \mathbb{E}_1[I(T_{A_{\langle r\rangle}} < T_0)L]$, where $\mathbb{E}_1[\cdot]$ denotes expectation under IS with L as the resulting likelihood ratio. Under $\mathbb{P}_0$, the distribution of $T_{A_{\langle r\rangle}}$ depends on r, but the distribution of T_0 does not. In (9), we apply MSIS using a random vector $\mathbf{X}$ of dimension $d = 2$, with $\mathbf{X} = (\min(T_{A_{\langle r\rangle}}, T_0), I(T_{A_{\langle r\rangle}} < T_0)L)$, and the expectation $\mathbb{E}[\cdot]$ applies $\mathbb{E}_0[\cdot]$ (resp., $\mathbb{E}_1$) to the first (resp., second) coordinate. Thus, $\boldsymbol{\theta}_{\langle r\rangle} = (\theta_{\langle r\rangle,1}, \theta_{\langle r\rangle,2}) = (\eta_{\langle r\rangle}, \gamma_{\langle r\rangle})$ and $g(x_1, x_1) = x_1/x_2$. For the first component of $\mathbf{X}$, $0 \le \min(T_{A_{\langle r\rangle}}, T_0) \le T_0$ for all r implies that $\mathbb{E}_0[\min(T_{A_{\langle r\rangle}}, T_0)] \le \mathbb{E}_0[T_0]$, with $\mathbb{E}_0[T_0] < \infty$ by, e.g., Corollary 1(c) of Thorisson (1985). Moreover, because $\mathbb{P}_0(\lim_{r\to\infty} T_{A_{\langle r\rangle}} = \infty) = 1$ and $\mathbb{P}_0(T_0 < \infty) = 1$ guarantee $\mathbb{P}_0(\lim_{r\to\infty} \min(T_{A_{\langle r\rangle}}, T_0) = T_0) = 1$, the dominated convergence theorem then ensures that $\eta_{\langle r\rangle} = \mathbb{E}_0[\min(T_{A_{\langle r\rangle}}, T_0)] \to \mathbb{E}_0[T_0]$ as $r \to \infty$. In addition, $\chi_1 \equiv \mathbb{E}_0[T_0] \in (0, \infty)$ since $0 < \mathbb{E}_0[U] \le \mathbb{E}_0[T_0] < \infty$, so $\delta_1 = 0$ in (12). Also, Theorem 1 of Sadowsky (1991) shows that $\gamma_{\langle r\rangle} \to 0$ as $r \to \infty$, so $\delta_2 = -\delta$ in (12) for each $\delta > 0$. Hence, the estimand satisfies $\alpha_{\langle r\rangle} = \eta_{\langle r\rangle}/\gamma_{\langle r\rangle} \to \infty$ as $r \to \infty$, so $\delta_0 = \delta$ in (12). We now want to verify (13) for $i = 1, 2$. Note that $g_1(\boldsymbol{\theta}_{\langle r\rangle}) = 1/\theta_{\langle r\rangle,2} = 1/\gamma_{\langle r\rangle}$ and $g_2(\boldsymbol{\theta}_{\langle r\rangle}) = -\theta_{\langle r\rangle,1}/\theta^2_{\langle r\rangle,2} = -\eta_{\langle r\rangle}/\gamma^2_{\langle r\rangle}$. For $i = 1$, because $\delta_1 = 0$ and $\delta_0 = \delta$, (11) becomes

$$h_{1,1+\delta_1,1+\delta_0}(\boldsymbol{\theta}_{\langle r\rangle}) = \left| \frac{\theta^{1+\delta_1}_{\langle r\rangle,1} g_1(\boldsymbol{\theta}_{\langle r\rangle})}{g^{1+\delta_0}(\boldsymbol{\theta}_{\langle r\rangle})} \right| = \left| \frac{\eta_{\langle r\rangle}/\gamma_{\langle r\rangle}}{(\eta_{\langle r\rangle}/\gamma_{\langle r\rangle})^{1+\delta}} \right| = \frac{1}{\alpha^{\delta}_{\langle r\rangle}} \to 0 \quad \text{as } r \to \infty \tag{18}$$

for each $\delta > 0$. Also, $\mathbb{E}_0[(\min(T_{A_{\langle r\rangle}}, T_0))^2] \le \mathbb{E}_0[T_0^2]$ for all r, where $\mathbb{E}_0[T_0^2]$ does not depend on r and $\mathbb{E}_0[T_0^2] < \infty$ by Corollary 1(c) of Thorisson (1985) since $\mathbb{E}_0[U^2] < \infty$ and $\mathbb{E}_0[V^2] < \infty$. In this case, under $\mathbb{P}_0$, the variance of $\min(T_{A_{\langle r\rangle}}, T_0)$ is bounded in r, and we can then estimate $\theta_{\langle r\rangle,1} = \eta_{\langle r\rangle}$ with BRE using naive MC. Therefore, $\mathrm{RE}_{1+\delta_1}[\widehat{\theta}_{\langle r\rangle,n,1}] = \mathrm{RE}[\widehat{\theta}_{\langle r\rangle,n,1}] = O(1)$ as $r \to \infty$, which combined with (18) verifies (13) for $i = 1$. For $i = 2$, because $\delta_2 = -\delta$ and $\delta_0 = \delta$ in (12), (11) becomes

$$h_{2,1+\delta_2,1+\delta_0}(\boldsymbol{\theta}_{\langle r\rangle}) = \left| \frac{\theta^{1+\delta_2}_{\langle r\rangle,2} g_2(\boldsymbol{\theta}_{\langle r\rangle})}{g^{1+\delta_0}(\boldsymbol{\theta}_{\langle r\rangle})} \right| = \left| \frac{\gamma^{1-\delta}_{\langle r\rangle}(-\eta_{\langle r\rangle}/\gamma^2_{\langle r\rangle})}{(\eta_{\langle r\rangle}/\gamma_{\langle r\rangle})^{1+\delta}} \right| = \frac{1}{\eta^{\delta}_{\langle r\rangle}}, \tag{19}$$

which, for each $\delta > 0$, is bounded since $\eta_{\langle r \rangle} \geq \mathbb{E}_0[U] > 0$. Sadowsky (1991) provides conditions under which an IS estimator (via exponential twisting) for $\theta_{\langle r \rangle,2} = \gamma_{\langle r \rangle}$ is LE, so this estimator satisfies $\limsup_{r\to\infty} \mathrm{RE}_{1+\delta_2}[\widehat{\theta}_{\langle r \rangle,n,2}] = \limsup_{r\to\infty} \mathrm{RE}_{1-\delta}[\widehat{\theta}_{\langle r \rangle,n,2}] = 0$ as in (6), which combined with (19) confirms (13) for $i = 2$. Hence, the MSIS estimator of $\alpha_{\langle r \rangle}$ is LE by Theorem 1(iii).

3.2 Special Case When $d = 1$

We now examine Theorem 1 in more detail for the special case when the dimension $d = 1$, so in (8), the unknown parameter is $\boldsymbol{\theta}_{\langle r \rangle} = \theta_{\langle r \rangle,1} \equiv \theta_{\langle r \rangle}$, the estimand is $\alpha_{\langle r \rangle} = g(\theta_{\langle r \rangle})$, and the random vector $\mathbf{X}_{\langle r \rangle} = X_{\langle r \rangle,1} \equiv X_{\langle r \rangle}$ has mean $\theta_{\langle r \rangle}$. Theorem 1 then requires that the derivative g' of g is nonzero at $\theta_{\langle r \rangle}$, and the asymptotic variance in (16) becomes $\tau^2_{\langle r \rangle} = [g'(\theta_{\langle r \rangle})]^2 \mathrm{Var}[X_{\langle r \rangle}]$.

We first focus on BRE/VRE properties, which correspond to parts (i) and (ii) of Theorem 1. For $\widehat{\theta}_{\langle r \rangle,n}$ as the estimator of $\theta_{\langle r \rangle}$, the relative error of the estimator $g(\widehat{\theta}_{\langle r \rangle,n})$ of $g(\theta_{\langle r \rangle})$ is

$$\mathrm{RE}[g(\widehat{\theta}_{\langle r \rangle,n})] = \frac{|g'(\theta_{\langle r \rangle})|\sqrt{\mathrm{Var}[X_{\langle r \rangle}]}}{|g(\theta_{\langle r \rangle})|} = \left|\frac{\theta_{\langle r \rangle} g'(\theta_{\langle r \rangle})}{g(\theta_{\langle r \rangle})}\right| \frac{\sqrt{\mathrm{Var}[X_{\langle r \rangle}]}}{|\theta_{\langle r \rangle}|} = h(\theta_{\langle r \rangle}) \mathrm{RE}[\widehat{\theta}_{\langle r \rangle,n}], \tag{20}$$

where $h(\theta_{\langle r \rangle}) = |\theta_{\langle r \rangle} g'(\theta_{\langle r \rangle})/g(\theta_{\langle r \rangle})|$ and $\mathrm{RE}[\widehat{\theta}_{\langle r \rangle,n}] = \sqrt{\mathrm{Var}[X_{\langle r \rangle}]}/|\theta_{\langle r \rangle}|$. Thus, $g(\widehat{\theta}_{\langle r \rangle,n})$ has BRE (resp., VRE) if and only if $h(\theta_{\langle r \rangle})\mathrm{RE}[\widehat{\theta}_{\langle r \rangle,n}]$ remains bounded (resp., vanishes) as $\varepsilon \to 0$. As discussed just before the proof of Theorem 1, a sufficient condition for $g(\widehat{\theta}_{\langle r \rangle,n})$ to have BRE (resp., VRE) is that $\widehat{\theta}_{\langle r \rangle,n}$ has BRE (resp., VRE) and $h(\theta_{\langle r \rangle})$ remains bounded as $\varepsilon \to 0$. But $g(\widehat{\theta}_{\langle r \rangle,n})$ can still have BRE or VRE even when $\widehat{\theta}_{\langle r \rangle,n}$ does not if $h(\theta_{\langle r \rangle}) \to 0$ fast enough as $r \to \infty$. We next consider some specific examples.

First suppose that $g(\theta_{\langle r \rangle}) = c[\theta_{\langle r \rangle}]^v$ for constants $c \neq 0$ and $v \neq 0$. In this case, we get $h(\theta_{\langle r \rangle}) = |\theta_{\langle r \rangle} cv[\theta_{\langle r \rangle}]^{v-1}/(c[\theta_{\langle r \rangle}]^v)| = |v|$, so $h(\theta_{\langle r \rangle})$ is a constant and is trivially bounded, no matter how $\theta_{\langle r \rangle}$ depends on r. Thus, it then follows that $g(\widehat{\theta}_{\langle r \rangle,n})$ has BRE or VRE if and only if $\widehat{\theta}_{\langle r \rangle,n}$ does.

Now instead suppose $g(\theta_{\langle r \rangle}) = ce^{\theta_{\langle r \rangle}}$ for some constant $c \neq 0$, so $h(\theta_{\langle r \rangle}) = |\theta_{\langle r \rangle} ce^{\theta_{\langle r \rangle}}/(ce^{\theta_{\langle r \rangle}})| = |\theta_{\langle r \rangle}|$. If $\theta_{\langle r \rangle} = c_1 r^\ell$ for constants $c_1 \neq 0$ and $\ell > 0$, then $h(\theta_{\langle r \rangle}) = |c_1 r^\ell| \to \infty$ as $r \to \infty$. Thus, it is possible to have unbounded $h(\theta_{\langle r \rangle})$. Moreover, suppose that $\mathrm{Var}[X_{\langle r \rangle}] = c_2^2 \theta^2_{\langle r \rangle} + o(\theta^2_{\langle r \rangle})$ as $r \to \infty$ for some constant $c_2 > 0$, so that $\mathrm{RE}[\widehat{\theta}_{\langle r \rangle,n}] = c_2 + o(1)$, i.e., $\theta_{\langle r \rangle}$ has BRE (but not VRE); then (20) implies $\mathrm{RE}[g(\widehat{\theta}_{\langle r \rangle,n})] = h(\theta_{\langle r \rangle})\mathrm{RE}[\widehat{\theta}_{\langle r \rangle,n}] \to \infty$, showing that $g(\widehat{\theta}_{\langle r \rangle,n})$ can have unbounded RE, even though $\widehat{\theta}_{\langle r \rangle,n}$ has BRE. On the other hand, if $\theta_{\langle r \rangle} = c_1 r^{-\ell}$ for $\ell > 0$, then $h(\theta_{\langle r \rangle}) = |c_1 r^{-\ell}| \to 0$ as $r \to \infty$. In this case, if $\mathrm{RE}[\widehat{\theta}_{\langle r \rangle,n}] = O(r^\ell)$ (resp., $o(r^\ell)$), then $g(\widehat{\theta}_{\langle r \rangle,n})$ has BRE (resp., VRE), even though $\widehat{\theta}_{\langle r \rangle,n}$ may have unbounded RE.

For LE, focus now on part (iii) of Theorem 1. As in (20), with δ_0 and δ_1 defined in (12), we get

$$\mathrm{RE}_{1+\delta_0}[g(\widehat{\theta}_{\langle r \rangle,n})] = \frac{|g'(\theta_{\langle r \rangle})|\sqrt{\mathrm{Var}[X_{\langle r \rangle}]}}{|g(\theta_{\langle r \rangle})|^{1+\delta_0}} = \left|\frac{\theta^{1+\delta_1}_{\langle r \rangle} g'(\theta_{\langle r \rangle})}{(g(\theta_{\langle r \rangle}))^{1+\delta_0}}\right| \frac{\sqrt{\mathrm{Var}[X_{\langle r \rangle}]}}{|\theta_{\langle r \rangle}|^{1+\delta_1}} = h_{1+\delta_1,1+\delta_0}(\theta_{\langle r \rangle}) \mathrm{RE}_{1+\delta_1}[\widehat{\theta}_{\langle r \rangle,n}], \tag{21}$$

where $h_{1+\delta_1,1+\delta_0}(\theta_{\langle r \rangle}) = \left|\theta^{1+\delta_1}_{\langle r \rangle} g'(\theta_{\langle r \rangle})/(g(\theta_{\langle r \rangle}))^{1+\delta_0}\right|$. Here too, a sufficient condition for $g(\widehat{\theta}_{\langle r \rangle,n})$ to be LE is that $\widehat{\theta}_{\langle r \rangle,n}$ is LE and $h_{1+\delta_1,1+\delta_0}(\theta_{\langle r \rangle})$ remains bounded. Also if $h_{1+\delta_1,1+\delta_0}(\theta_{\langle r \rangle}) \to 0$ sufficiently fast, it may be possible for $g(\widehat{\theta}_{\langle r \rangle,n})$ to have BRE or VRE when $\widehat{\theta}_{\langle r \rangle,n}$ is only LE. For example, again considering $g(\theta_{\langle r \rangle}) = ce^{\theta_{\langle r \rangle}}$ for some constant $c \neq 0$ and assuming $\theta_{\langle r \rangle} \to 0$ as $r \to \infty$ (so that $\delta_1 = -\delta$), we get $h_{1+\delta_1,1+\delta_0}(\theta_{\langle r \rangle}) = |\theta_{\langle r \rangle}|^{1-\delta}/|g(\theta_{\langle r \rangle})|^{\delta_0}$. Recall that (5) with $\zeta = 1$ reduces to (2), so using $\delta_0 = 0$ in

(21) yields $\mathrm{RE}[g(\widehat{\theta}_{\langle r\rangle,n})] = |\theta_{\langle r\rangle}|^{1-\delta}\mathrm{RE}_{1-\delta}[\widehat{\theta}_{\langle r\rangle,n}] \to 0$ as $r\to\infty$ by (6) since we assumed that $\widehat{\theta}_{\langle r\rangle,n}$ is LE; thus, $g(\widehat{\theta}_{\langle r\rangle,n})$ has VRE even though $\widehat{\theta}_{\langle r\rangle,n}$ is only LE.

4 NUMERICAL RESULTS

Example 9 Nakayama and Tuffin (2022) describe a reliability system with two components, which we now consider. The first component fails after an exponential amount of time, and then with probability $1-p$, it automatically resets, and the process repeats again independently of the past. If the first component does not reset, which occurs with probability p, the second component takes over, and has a uniformly distributed lifetime. After the second component fails, the system fails, and then the entire system resets after an exponential amount of time. We model the system as a 3-state semi-Markov process (SMP) with state space $S=\{0,1,2\}$, whose embedded DTMC has transition probabilities as in Figure 1. The figure also shows the holding-time distribution for each successive visit to a state $s\in S$, where $\mathscr{E}(\lambda)$ denotes an exponential distribution with rate $\lambda>0$, so its mean is $1/\lambda$, and $\mathscr{U}(a,b)$ is a uniform distribution on the interval (a,b). We define a rarity parameter $\varepsilon=1/r$, and let w_0, w_1, and λ_2 be positive constants, whose values we will vary to study different behaviors of the model as $r\to\infty$.

We aim at computing the mean hitting time of state 2, which $\alpha_{\langle r\rangle}$ now represents. Typically, as explained in Nakayama and Tuffin (2022), if $w_0+1>w_1$, the time spent in state 1 will make a negligible contribution to the hitting time compared to the total time spent in state 0 before moving to state 1, and both the estimator based on a ratio as in Example 5 or the alternative estimator of Example 6 will be efficient. In contrast, if $w_0+1\le w_1$, the estimator in Example 5 encounters issues, motivating the approach in Example 6.

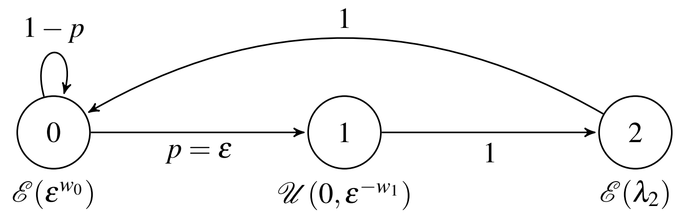

Figure 1: The edge labels are the transition probabilities of the embedded DTMC of an SMP, with the holding-time distribution to each visit to a state below that state, where w_0, w_1, and λ_2 are constants.

With the notation in Example 6, the rare set is $A=\{2\}$, $T_{0,\langle r\rangle}$ is the time to first return to state 0, and $T_{A,\langle r\rangle}$ is the system failure time. Our experiments used $w_0=1$ and $w_1=3$, and we considered several values of $\varepsilon=1/r$, simulating with MSIS a total number $n=10^6$ of cycles, where a (regenerative) cycle is a sample path between two successive entrances to state 0. While CMC may also be applied by conditioning on the SMP's embedded DTMC, we did not do this.) For MSIS, we generate a number $n_1=\lfloor\kappa n\rfloor$ (with $0<\kappa<1$) of cycles to estimate $\theta_{\langle r\rangle,1}=\mathbb{E}_0[\min(T_{2,\langle r\rangle},T_{0,\langle r\rangle})I(T_{2,\langle r\rangle}>T_{0,\langle r\rangle})]$ by naive MC, while the remaining $n_2=n-\lfloor\kappa n\rfloor$ cycles are used to estimate $\theta_{\langle r\rangle,2}=\mathbb{E}_1[\min(T_{2,\langle r\rangle},T_{0,\langle r\rangle})I(T_{2,\langle r\rangle}<T_{0,\langle r\rangle})L]$ and $\theta_{\langle r\rangle,3}=\mathbb{E}_1[I(T_{2,\langle r\rangle}<T_{0,\langle r\rangle})L]$ by IS, which replaces the transition probability $p=\varepsilon=1/r$ from state 0 to state 1 by a probability 0.8. The proportion κ is computed from a presimulation to minimize the work-normalized variance of the resulting estimator (its expression is denoted γ^* in Equation (23) of Nakayama and Tuffin 2019), but keeping at least 10% of the cycles to both naive and IS simulations. This potentially unequal sampling allocation using κ differs from the setting in (10), which assumes that each θ_i is estimated with the same sample size (but possibly different simulation methods), but this does not affect the asymptotic behavior (ignoring leading constants) since we constrain $\kappa\in(0.1,0.9)$.

Table 1 displays the output for the estimation of $\widehat{\theta}_{\langle r\rangle,n_i,i}$ of $\theta_{\langle r\rangle,i}$ ($i\in\{1,2,3\}$, with $n_3=n_2$), an estimate $\widehat{\mathrm{RE}}^{(i)}$ of $\mathrm{RE}[\widehat{\theta}_{\langle r\rangle,n_i,i}]$, and the resulting estimation $\widehat{\alpha}_{\langle r\rangle,n}$ of the expected failure time with the corresponding

Table 1: Results for the MTTF in the SMP example of Figure 1, from $n = 10^6$ independent cycles.

$\varepsilon = 1/r$	$\widehat{\theta}_{\langle r\rangle,n_1,1}$	$\widehat{\mathrm{RE}}^{(1)}$	$\widehat{\theta}_{\langle r\rangle,n_2,2}$	$\widehat{\mathrm{RE}}^{(2)}$	$\widehat{\theta}_{\langle r\rangle,n_2,3}$	$\widehat{\mathrm{RE}}^{(3)}$	$\widehat{\alpha}_{\langle r\rangle,n}$	$\widehat{\mathrm{RE}}^{(0)}$	95% CI
1.0e-01	9.0e+00	1.1e+00	5.1e+01	8.1e-01	1.0e-01	5.0e-01	6.0e+02	7.1e-01	(6.0e+02, 6.0e+02)
1.0e-02	9.9e+01	1.0e+00	5.0e+03	8.2e-01	1.0e-02	5.0e-01	5.1e+05	6.7e-01	(5.1e+05, 5.1e+05)
1.0e-03	1.0e+03	1.0e+00	5.0e+05	8.2e-01	1.0e-03	5.0e-01	5.0e+08	6.8e-01	(5.0e+08, 5.0e+08)
1.0e-04	1.0e+04	1.0e+00	5.0e+07	8.2e-01	1.0e-04	5.0e-01	5.0e+11	6.8e-01	(5.0e+11, 5.0e+11)

estimate $\widehat{\mathrm{RE}}^{(0)}$ of $\mathrm{RE}[\widehat{\alpha}_{\langle r\rangle,n}]$ and CI. The results appear to show that the RE of the estimator of each θ_i does not grow as ε shrinks, and the same is true when estimating $\alpha_{\langle r\rangle} = \mathbb{E}[T_{A,\langle r\rangle}]$. This behavior indicates BRE, which agrees with the theory in Theorem 1 and Example 6.

Example 10 We consider an HRMS (as in Example 5) with $c = 3$ types of component and a redundancy 5 for each type. Each component has an exponentially distributed time to failure with rate $\varepsilon = 1/r$. Any failed component has an exponentially distributed repair time with rate 1, where there are 15 repairpersons, so no failed component ever has to queue before receiving service. Component failure and repair times are all independent. The system is down whenever fewer than two components of any one type are operational. We want to compute here the *steady-state unavailability*, so $\alpha_{\langle r\rangle}$ now represents the long run fraction of time that the resulting CTMC is down. As in Goyal et al. (1992), we can express $\alpha_{\langle r\rangle} = g(\theta_{\langle r\rangle,1}, \theta_{\langle r\rangle,2}) = \theta_{\langle r\rangle,1}/\theta_{\langle r\rangle,2}$, where the numerator is the expected down time in a cycle, and the denominator is the expected cycle length, where a cycle is a sample path between two successive entrances to the fully operational state s_0. Applying CMC by conditioning on the embedded DTMC, we can write $\theta_{\langle r\rangle,1}$ (resp., $\theta_{\langle r\rangle,2}$) as the mean of the sum of the expected sojourn times of each failed state (resp., all states) visited within a cycle. We consider MSIS to estimate $\theta_{\langle r\rangle,1}$ and $\theta_{\langle r\rangle,2}$ from a total number $n = 10^6$ of independent cycles of the embedded DTMC, with $n_1 = \lfloor \kappa n \rfloor$ (resp. $n_2 = n - n_1$) cycles used to estimate $\theta_{\langle r\rangle,1}$ (resp. $\theta_{\langle r\rangle,2}$), with κ obtained (as in Example 9) from a presimulation to minimize the work-normalized variance of the resulting estimator. We estimate $\theta_{\langle r\rangle,1}$ through an approach known as *dynamic IS* (Goyal et al. 1992), which within a cycle applies IS only up to the first time the failed set is reached, at which point the rest of the cycle reverts to naive MC. The IS employs *balanced failure biasing* (Shahabuddin 1994), which, in a state from which both failures and repairs are possible, assigns a cumulative probability $\rho = 0.8$ to all failure transitions, equally allocating ρ across the individual failure transitions; repair transitions get a total probability of $1 - \rho$, distributed to individual repair transitions in proportion to their original probabilities. Shahabuddin (1994) establishes that this dynamic IS procedure yields an estimator of $\theta_{\langle r\rangle,1}$ with BRE. We utilize naive simulation to estimate $\theta_{\langle r\rangle,2}$. Table 2 illustrates again that BRE is satisfied for all estimators.

Table 2: Results for unavailability for the HRMS of Example 10, from $n = 10^6$ independent cycles.

ε	$\widehat{\theta}_{\langle r\rangle,n_1,1}$	$\widehat{\mathrm{RE}}^{(1)}$	$\widehat{\theta}_{\langle r\rangle,n_2,2}$	$\widehat{\mathrm{RE}}^{(2)}$	$\widehat{\alpha}_{\langle r\rangle,n}$	$\widehat{\mathrm{RE}}[\widehat{\alpha}_{\langle r\rangle,n}]$	95% CI
1.00e-01	2.90e-03	8.85e+01	2.78e+00	8.85e-01	1.04e-03	9.33e+01	(8.516e-04, 1.233e-03)
1.00e-02	1.11e-06	5.62e+00	7.74e+00	7.29e-02	1.43e-07	5.93e+00	(1.414e-07, 1.447e-07)
1.00e-03	1.01e-09	6.98e+00	6.77e+01	2.55e-03	1.49e-11	7.35e+00	(1.467e-11, 1.510e-11)
1.00e-04	1.00e-12	7.17e+00	6.68e+02	8.45e-05	1.50e-15	7.56e+00	(1.476e-15, 1.521e-15)
1.00e-05	9.98e-16	7.20e+00	6.67e+03	3.46e-06	1.50e-19	7.59e+00	(1.475e-19, 1.519e-19)
1.00e-06	9.98e-19	7.20e+00	6.67e+04	1.16e-06	1.50e-23	7.59e+00	(1.475e-23, 1.519e-23)

The literature provides many other numerical examples displaying empirical behavior consistent with the theory in Theorem 1. For the MTTF ratio estimator in Example 5, Goyal et al. (1992) show the BRE property through numerical results on a model of a computing system. When estimating the mean time for the queue-length process to hit a large buffer size r in a stable M/M/1 queue, Nakayama and Tuffin (2019) present simulation results using the ratio estimator in Examples 2 and 8 and the alternative estimator described in Example 6, with the latter approach outperforming the former for heavier congestion levels. The M/M/1 estimators have LE but not BRE.

5 CONCLUDING REMARKS

We considered estimating a function of a vector of means in a rare-event context. This covers a wide range of estimands that arise in practice, as we illustrated through numerous examples. Theorem 1 provides conditions that show when the overall estimator can be estimated efficiently in terms of the efficiency behaviors of the estimators of each individual mean.

REFERENCES

Asmussen, S., and P. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*. New York: Springer.

Billingsley, P. 1995. *Probability and Measure*. 3rd ed. New York: John Wiley and Sons.

Glynn, P. W. 1990. "Likelihood Ratio Derivative Estimators for Stochastic Systems". *Communications of the ACM* 33:75–84.

Glynn, P. W., M. K. Nakayama, and B. Tuffin. 2017. "On the Estimation of the Mean Time to Failure by Simulation". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, A.D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 1844–1855. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Goyal, A., P. Shahabuddin, P. Heidelberger, V. Nicola, and P. W. Glynn. 1992. "A Unified Framework for Simulating Markovian Models of Highly Dependable Systems". *IEEE Transactions on Computers* C-41(1):36–51.

Kalashnikov, V. 1994. *Topics on Regenerative Processes*. Boca Raton: CRC Press.

L'Ecuyer, P., J. H. Blanchet, B. Tuffin, and P. W. Glynn. 2010. "Asymptotic Robustness of Estimators in Rare-Event Simulation". *ACM Transactions on Modeling and Computer Simulation* 20(1):Article 6.

L'Ecuyer, P., and B. Tuffin. 2012. "Approximating Zero-Variance Importance Sampling in a Reliability Setting". *Annals of Operations Research* 189(1):277–297.

Nakayama, M. K. 1996. "General Conditions for Bounded Relative Error in Simulations of Highly Reliable Markovian Systems". *Advances in Applied Probability* 28:687–727.

Nakayama, M. K. 1998. "On Derivative Estimation of the Mean Time to Failure in Simulations of Highly Reliable Markovian Systems". *Operations Research* 46:285–290.

Nakayama, M. K., and B. Tuffin. 2019. "Efficient Estimation of the Mean Hitting Time to a Set of a Regenerative System". In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 416–427. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Nakayama, M. K., and B. Tuffin. 2022. "Density Estimators of the Cumulative Reward up to a Hitting Time to a Rarely Visited Set of a Regenerative System". In *Proceedings of the 2022 Winter Simulation Conference*, edited by B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C. Corlu, L. Lee, E. Chew, T. Roeder, and P. Lendermann, 121–132. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Rubino, G., and B. Tuffin. (Eds.) 2009. *Rare Event Simulation using Monte Carlo Methods*. Chichester, UK: John Wiley.

Sadowsky, J. S. 1991. "Large Deviations Theory and Efficient Simulation of Excessive Backlogs in a GI/GI/m Queue". *IEEE Transactions on Automatic Control* 36:1383–1394.

Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics*. New York: John Wiley and Sons.

Shahabuddin, P. 1994. "Importance Sampling for Highly Reliable Markovian Systems". *Management Science* 40:333–352.

Shao, J., and D. Tu. 1995. *The Jackknife and Bootstrap*. New York: Springer.

Thorisson, H. 1985, February. "The Queue GI/G/1: Finite Moments of the Cycle Variables and Uniform Rates of Convergence". *Stochastic Processes and their Applications* 19(1):85–99.

AUTHOR BIOGRAPHIES

MARVIN K. NAKAYAMA is a professor in the Department of Computer Science at the New Jersey Institute of Technology. He received a Ph.D. in operations research from Stanford University, and is an associate editor for *ACM Transactions on Modeling and Computer Simulation*. His email: marvin@njit.edu.

BRUNO TUFFIN received his PhD degree in applied mathematics from the University of Rennes 1, France, in 1997. Since then, he has been with Inria in Rennes. His research interests include developing Monte Carlo and quasi-Monte Carlo simulation techniques for the performance evaluation of telecommunication systems and telecommunication-related economical models. He is currently Area Editor for *INFORMS Journal on Computing* and Associate Editor for *ACM Transactions on Modeling and Computer Simulation* and *Queuing Systems*. His email: bruno.tuffin@inria.fr.