Robustness Properties for Simulations of Highly Reliable Systems

Bruno Tuffin[*], Pierre L'Ecuyer[†], and Werner Sandmann[‡]

Abstract

Importance sampling (IS) is the primary technique for constructing reliable estimators in the context of rare-event simulation. The asymptotic robustness of IS estimators is often qualified by properties such as bounded relative error (BRE) and asymptotic optimality (AO). These properties guarantee that the estimator's relative error remains bounded (or does not increase too fast) when the rare events becomes rarer. Other recently introduced characterizations of IS estimators are bounded normal approximation (BNA), bounded relative efficiency (BREff), and asymptotic good estimation of mean and variance.

In this paper we introduce three additional property named bounded relative error of empirical variance (BREEV), bounded relative efficiency of empirical variance (BREffEV), and asymptotic optimality of empirical variance (AOEV), which state that the empirical variance has itself the BRE, BREff and AO property, respectively, as an estimator of the true variance. We then study the hierarchy between all these different characterizations for a model of highly-reliable Markovian systems (HRMS) where the goal is to estimate the failure probability of the system. In this setting, we show that BRE, BREff and AO are equivalent, that BREffEV, BREEV and AOEV are also equivalent, and that these two properties are strictly stronger than all other properties just mentioned. We also obtain a necessary and sufficient condition for BREEV in terms of quantities that can be readily verified from the parameters of the model.

1 Introduction

Rare event simulation has received a lot of attention due to its frequent occurrence in areas such as reliability, telecommunications, finance, and insurance, among others [3, 11, 12]. In typical rare-event settings, Monte Carlo simulation is not viable unless special "acceleration" techniques are used to make the important rare events occur frequently enough for moderate sample sizes. The two main families of techniques for doing that are splitting [8, 13, 22] and importance sampling (IS) [3, 9, 11].

Asymptotic analysis of rare-event simulations is usually made in an asymptotic regime where rarity is controlled by a parameter $\varepsilon > 0$; the rare events become increasingly rare when $\varepsilon \to 0$ and we are interested in asymptotic properties of a given (unbiased) estimator Y in the limit. (Some authors use a parameter m that goes to infinity instead, but this is equivalent; it suffice to take $\varepsilon = 1/m$ to recover our framework.) Asymptotic characterizations of estimators in this setting include the widely-used concepts of *bounded relative error* (BRE) and *asymptotic optimality* (AO) [11, 12], as well as the lesser-known properties of *bounded relative efficiency* (BREff) [5], *bounded normal approximation* (BNA) and *asymptotic good estimation of the mean* (AGEM) and *of the variance* (AGEV) (also called probability and variance well-estimation) [19, 20].

BRE means that the relative error (the standard deviation divided by the mean) of the estimator $Y = Y(\varepsilon)$ remains bounded when $\varepsilon \to 0$. AO requires that when the mean converges to zero exponentially fast in ε, the standard deviation converges at the same exponential rate. In general, this is a weaker condition than BRE [11, 16]. BREff generalizes BRE by taking into account the computational time associated with the

---

[*]IRISA/INRIA, Campus de Beaulieu, 35042 Rennes Cedex, France, btuffin@irisa.fr

[†]DIRO, Université de Montreal, C.P. 6128, Succ. Centre-Ville, Montréal (Québec), Canada, H3C 3J7, lecuyer@iro.umontreal.ca

[‡]Otto-Friedrich-Universität Bamberg, Feldkirchenstr. 21, D-96045 Bamberg, Germany, werner.sandmann@wiai.uni-bamberg.de

estimator Y, which may vary with ε. BNA implies that if we approximate the distribution of the average of n i.i.d. copies of Y by the normal distribution (e.g., to compute a confidence interval), the quality of the approximation does not degrade when $\varepsilon \to 0$. AGEM and AGEV have been defined in the context of estimating a probability in a HRMS, and basically mean that the sample paths that contribute the most to the estimator and its second moment are not rare under the sampling scheme that is examined. The main goals of splitting and IS, from the asymptotic viewpoint, is to design estimators that enjoy some (or all) of these properties when the original (crude or naive) estimator does not satisfy them.

An important difficulty often lurking around in rare-event simulation is that of estimating the variance of the mean estimator: reliable variance estimators are typically more difficult to obtain than reliable mean estimators, because the rare events have a stronger influence on the variance than on the mean. Variance estimators are important because we need them to assess the accuracy of our mean estimators, e.g., via confidence intervals. They are also frequently used when we compare the efficiencies of alternative mean estimators; poor variance estimators can easily yield misleading results in this context. This motivates our introduction of three additional characterizations of estimators: *bounded relative error of empirical variance* (BREEV), *bounded relative efficiency of empirical variance* (BREffEV), and *asymptotically optimal empirical variance* (AOEV). BREEV means that the empirical variance has the BRE property while AOEV means that it has the AO property.

In this paper, we focus on IS and its application to an important HRMS model studied by several authors [4, 10, 11, 14, 15, 17, 19, 20], and used for reliability analysis of computer and telecommunication systems. In this model, a smaller value of the rarity parameter ε implies a smaller failure rate for the system's components, and we want to estimate the probability that the system reaches a "failed" state before it returns to a state where all the components are operational. This probability converges to 0 when $\varepsilon \to 0$.

In general, IS consists in simulating the original model with different (carefully selected) probability laws for its input random variables, and counter-balancing the bias caused by this change of measure with a weight called the likelihood ratio. For the HRMS model, we actually simulate a discrete-time Markov chain whose transitions correspond to failures and repairs of individual components and IS generally increases [decreases] the probabilities of the failure [repair] transitions.

For this particular HRMS model, specific conditions on the model parameters and IS probabilities have been obtained for the BRE property [15], for BNA [19, 20], and for AGEM and AGEV [20]. It is also shown in [20] that BNA implies AGEV, which implies BRE, which implies AGEM, which implies BRE, and that for each implication the converse is not true. In this paper we extend this hierarchy to incorporate AO, BREEV, and AOEV. We show that in our context, BRE, BREff and AO are equivalent, BREEV, BREffEV and AOEV are equivalent, and the latter three properties are strictly stronger than all the others. We also obtain a necessary and sufficient condition on the model parameters and the IS measure for BREEV, BREffEV and AOEV to hold.

The remainder of the paper is organized as follows. In Section 2, we give formal definitions of the asymptotic characterizations discussed so far: BRE, AO, BREff, BNA, AGEV, AGEM, BREEV, BREffEV, and AOEV, in a general rare-event framework. In Section 3, we recall the basic definition of IS in its general form. In Section 4, we describe the HRMS model and how IS is applied to this model. Section 5 is devoted studying to asymptotic robustness properties in the HRMS context. We establish a complete hierarchy between these properties and derive easily verifiable conditions for BREEV, BREffEV and AOEV. Finally, in Section 6, we conclude and highlight perspectives for further research.

The following notation is used all along the paper. For a function $f : (0, \infty) \to \mathbb{R}$, we say that $f(\varepsilon) = o(\varepsilon^d)$ if $f(\varepsilon)/\varepsilon^d \to 0$ as $\varepsilon \to 0$; $f(\varepsilon) = O(\varepsilon^d)$ if $|f(\varepsilon)| \le c_1 \varepsilon^d$ for some constant $c_1 > 0$ for all ε sufficiently small; $f(\varepsilon) = \underline{O}(\varepsilon^d)$ if $|f(\varepsilon)| \ge c_2 \varepsilon^d$ for some constant $c_2 > 0$ for all ε sufficiently small; and $f(\varepsilon) = \Theta(\varepsilon^d)$ if $f(\varepsilon) = \underline{O}(\varepsilon^d)$ and $f(\varepsilon) = O(\varepsilon^d)$.

2   Asymptotic Robustness Properties in a Rare-Event Setting

Rare-event framework.   We want to estimate a positive value $\gamma = \gamma(\varepsilon)$ that depends on a *rarity* parameter $\varepsilon > 0$. We assume that γ is a monotone (strictly) increasing function of ε and that $\lim_{\varepsilon \to 0^+} \gamma(\varepsilon) = 0$.

We have at our disposal a family of estimators $Y = Y(\varepsilon)$ such that $\mathbb{E}[Y(\varepsilon)] = \gamma(\varepsilon)$ for each $\varepsilon > 0$. Recall that the *variance* and *relative error* of $Y(\varepsilon)$ are defined by

$$\sigma^2(\epsilon) = \text{Var}[Y(\varepsilon)] = \mathbb{E}[(Y(\varepsilon) - \gamma(\varepsilon))^2]$$

and

$$\text{RE}[Y(\varepsilon)] = (\text{Var}[Y(\varepsilon)])^{1/2}/\gamma(\varepsilon).$$

In applications, $\gamma(\varepsilon)$ is usually a performance measure in the model, defined as a mathematical expectation, and some model parameters are defined as functions of ε. For example, in queuing systems, the service time and inter-arrival time distributions and the buffer sizes might depend on ε, while in Markovian reliability models, the failure rates and repair rates might be functions of ε. The convergence $\gamma(\varepsilon) \to 0$ can be exponential, polynomial, etc.; this depend on the application and how the model is parameterized. Note that in all cases, the limit as $\gamma(\varepsilon) \to 0^+$ is the same as the limit as $\varepsilon \to 0^+$, because of the strict monotonicity.

We now define several properties that the family of estimators $\{Y(\varepsilon),\ \varepsilon > 0\}$ can have. In these definitions (and elsewhere) we use the shorthand notation $Y(\varepsilon)$ to refer to this family (a slight abuse of notation). We write "$\to 0$" to mean "$\to 0^+$." In typical rare-event settings, these properties do not hold for the naive Monte Carlo estimators and the aim is to construct alternative unbiased estimators (e.g., via IS or other methods) for which they hold.

Bounded relative error.

Definition 1 *(BRE) The estimator $Y(\varepsilon)$ has the BRE property if*

$$\limsup_{\varepsilon \to 0} \text{RE}[Y(\epsilon)] < \infty. \tag{1}$$

When computing a confidence interval on $\gamma(\varepsilon)$ based on i.i.d. replications on $Y(\varepsilon)$ and the (classical) central-limit theorem, for a fixed confidence level, the width of the confidence interval is (approximately) proportional to the standard deviation $\sigma(\varepsilon)$. The BRE property means that this width decreases at least as fast as $\gamma(\varepsilon)$ when $\varepsilon \to 0$.

Asymptotic optimality. For several rare-event applications where $\gamma(\varepsilon)$ decreases exponentially fast (e.g., in queueing and finance), it has not been possible to find practical BRE estimators so far, but estimators with the (weaker) AO property have been constructed by exploiting the theory of large deviations [1, 7, 11, 12, 18]. AO means that when $\gamma^2(\varepsilon)$ converges to zero exponentially fast, the second moment $\mathbb{E}[Y^2(\varepsilon)]$ also converges exponentially fast and at the same exponential rate. This is the best possible rate; it cannot converge at a faster rate because we always have $\mathbb{E}[Y^2(\varepsilon)] - \gamma^2(\varepsilon) = \sigma^2(\varepsilon) \geq 0$.

Definition 2 *(AO) The estimator $Y(\varepsilon)$ is AO if*

$$\lim_{\varepsilon \to 0} \frac{\ln \mathbb{E}[Y^2(\varepsilon)]}{\ln \gamma(\varepsilon)} = 2. \tag{2}$$

AO is generally weaker than BRE [11, 16]. But there are situations where the two are equivalent; this is what will happen in our HRMS setup in Section 4. The following examples illustrate the two possibilities.

Example 1 Suppose that $\gamma(\varepsilon) = \exp[-k/\epsilon]$ for some constant k and that our estimator has $\sigma^2(\varepsilon) = q(1/\varepsilon)\exp[-2k/\varepsilon]$ for some polynomial function q. Then, the AO property is easily verified, whereas BRE does not hold because $\text{RE}^2[Y(\varepsilon)] = q(1/\varepsilon) \to \infty$ when $\varepsilon \to 0$.

Example 2 Suppose now that $\gamma^2(\varepsilon) = q_1(\varepsilon) = \varepsilon^{t_1} + o(\varepsilon^{t_1})$ and $\mathbb{E}[Y^2(\varepsilon)] = q_2(\varepsilon) = \varepsilon^{t_2} + o(\varepsilon^{t_2})$. That is, both converge to 0 at a polynomial rate. Clearly, $t_2 \leq t_1$, because $\mathbb{E}[Y^2(\varepsilon)] - \gamma^2(\varepsilon) \geq 0$. We

3

have BRE if and only if (iff) $q_2(\varepsilon)/q_1(\varepsilon)$ remains bounded when $\varepsilon \to 0$, iff $t_2 = t_1$. On the other hand, $-\ln q_1(\varepsilon) = -\ln(\varepsilon^{t_1}(1 + o(1))) = -t_1 \ln(\varepsilon) - \ln(1 + o(1))$ and similarly for $q_2(\varepsilon)$ and t_2. Then,

$$\lim_{\varepsilon \to 0} \frac{\ln \mathbb{E}[Y^2(\varepsilon)]}{\ln \gamma(\varepsilon)} = \lim_{\varepsilon \to 0} \frac{t_2 \ln \varepsilon}{(t_1/2) \ln \varepsilon} = \frac{2t_2}{t_1}.$$

Thus, AO holds iff $t_2 = t_1$, which means that BRE and AO are equivalent in this case.

Bounded relative efficiency.

Definition 3 *(BREff) Let $t(\varepsilon)$ be the expected computational time to generate the estimator $Y(\varepsilon)$, whose variance is $\sigma^2(\varepsilon)$. The relative efficiency of $Y(\varepsilon)$ is defined by*

$$\mathrm{REff}[Y(\varepsilon)] = \frac{\gamma(\varepsilon)^2}{\sigma^2(\varepsilon) t(\varepsilon)} = \frac{1}{\mathrm{RE}^2[Y(\varepsilon)] t(\varepsilon)}.$$

We will say that $Y(\varepsilon)$ has bounded relative efficiency (BREff) if $\liminf_{\varepsilon \to 0} \mathrm{REff}[Y(\varepsilon)] > 0$.

BREff basically looks at the BRE property, but for a given computational budget. Indeed, the computation time may vary with ε; this has to be encompassed in the BRE property.

Example 3 *If $t(\varepsilon) = \Theta(1)$, then BREff and BRE are equivalent properties.*

Example 4 *In [5], an example with BREff but without BRE is exhibited. In that example, $t(\varepsilon) = O(\varepsilon)$ but $\mathrm{RE}[Y(\varepsilon)] = \underline{O}(\varepsilon^{-1})$. Conversely, we might have examples such that $t(\varepsilon) = \underline{O}(\varepsilon^{-1})$ and $\mathrm{RE}[Y(\varepsilon)] = \Theta(1)$ so that BRE is verified, but not BREff.*

Bounded normal approximation.

We mentioned earlier the computation of a confidence interval on $\gamma(\varepsilon)$ based on the central-limit theorem. This type of confidence interval is reliable if the sample average has approximately the normal distribution, so it is relevant to examine the quality of this normal approximation when $\varepsilon \to 0$. An error bound for this approximation is provided by the following version of the Berry-Esseen theorem [2]:

Theorem 1 *(Berry-Esseen) Let $Y_1, \ldots, Y_n$ be i.i.d. random variables with mean 0, variance σ^2, and third absolute moment $\beta_3 = \mathbb{E}[|Y_1|^3]$. Let $\bar{Y}_n$ and S_n^2 be the empirical mean and variance of $Y_1, \ldots, Y_n$, and let F_n denote the distribution function of the standardized sum (or Student statistic)*

$$S_n^* = \sqrt{n}\bar{Y}_n/S_n.$$

Then, there is an absolute constant $a < \infty$ such that for all $x \in \mathbb{R}$ and all $n \geq 2$,

$$|F_n(x) - \Phi(x)| \leq \frac{a\beta_3}{\sigma^3 \sqrt{n}},$$

where Φ is the standard normal distribution function. The classical result usually has σ in place of S_n in the definition of S_n^* [6]; in that case one can take $a = 0.8$ [21].

This result motivated the introduction of the BNA property in [19], which requires that the Berry-Esseen bound remains $O(n^{-1/2})$ when $\varepsilon \to 0$.

Definition 4 *(BNA) The estimator $Y(\varepsilon)$ is said to have the BNA property if*

$$\limsup_{\varepsilon \to 0} \frac{\mathbb{E}\left[|Y(\varepsilon) - \gamma(\varepsilon)|^3\right]}{\sigma^3(\varepsilon)} < \infty. \tag{3}$$

The BNA property *implies* that $\sqrt{n}|F_n(x) - \Phi(x)|$ remains bounded as a function of ε, i.e., that the approximation error of F_n by the normal distribution remains in $O(n^{-1/2})$. The reverse is not necessarily true, however. Perhaps it could seem more natural to *define* the BNA property as meaning that $\sqrt{n}|F_n(x) - \Phi(x)|$ remains bounded, but we keep Definition 4 because it has already been adopted in several papers and because it is often easier to obtain necessary and sufficient conditions for BNA with this definition.

If a confidence interval of level $1 - \alpha$ is obtained using the normal distribution while the true distribution is F_n, the error of coverage of the computed confidence interval does not exceed $2 \sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)|$. If that confidence interval is computed from an i.i.d. sample $Y_1(\varepsilon), \ldots, Y_n(\varepsilon)$ of $Y(\varepsilon)$, BNA implies that the coverage error remains in $O(n^{-1/2})$ when $\varepsilon \to 0$, with a hidden constant that does not depend on ε, so it is controlled.

Bounded relative error, bounded relative efficiency, and asymptotic optimality of the empirical variance. The next properties concern the stability of the empirical variance as an estimator of the true variance $\sigma^2(\varepsilon)$. Let $Y_1(\varepsilon), \ldots, Y_n(\varepsilon)$ be an i.i.d. sample of $Y(\varepsilon)$, where $n \geq 2$. The empirical mean and empirical variance are $\bar{Y}_n(\varepsilon) = (Y_1(\varepsilon) + \cdots + Y_n(\varepsilon))/n$ and

$$S_n^2 = S_n^2(\varepsilon) = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i(\varepsilon) - \bar{Y}_n(\varepsilon))^2.$$

When the variance and/or the relative error of an estimator are estimated by simulation in a rare-event setting, it happens frequently that $S_n^2(\varepsilon)$ takes a very small value (orders of magnitude smaller than the true variance, because the important rare events did not happen) with large probability $1 - p(\varepsilon)$, and an extremely large value with very small probability $p(\varepsilon)$, where $p(\varepsilon) \to 0$ when $\varepsilon \to 0$. This gross underestimation of the variance leads to wrong conclusions on the accuracy of the simulation, with high probability. This motivates the following definition.

Definition 5 *(BREEV and AOEV) The estimator $Y(\varepsilon)$ has the BREEV property if*

$$\limsup_{\varepsilon \to 0} \mathrm{RE}[S_n^2(\epsilon)] < \infty. \tag{4}$$

It has BREffEV property if

$$\liminf_{\varepsilon \to 0} \mathrm{REff}[S_n^2(\epsilon)] > 0. \tag{5}$$

It has the AOEV property if

$$\lim_{\varepsilon \to 0} \frac{\ln \mathbb{E}[S_n^4(\varepsilon)]}{\ln \sigma^2(\varepsilon)} = 2. \tag{6}$$

A classical result states that

$$\mathrm{Var}[S_n^2] = \frac{1}{n}\left(E[(Y(\varepsilon) - E[Y(\varepsilon)])^4] - \frac{n-3}{n-1}\sigma^4 \right). \tag{7}$$

Thus, the BREEV, BREffEV, and AOEV properties are linked with the fourth moment of $Y(\varepsilon)$.

Asymptotic good estimation of the mean and of the variance. AGEM and AGEV are two additional robustness properties introduced in [20], under the name of "well estimated mean and variance," in the context of the application of IS to an HRMS model. Here we provide more general definitions of these properties. We assume that $Y(\varepsilon)$ is a *discrete* random variable, which takes value y with probability $p(\varepsilon, y) = \mathbb{P}[Y(\varepsilon) = y]$, for $y \in \mathbb{R}$. We also assume that its mean and variance are polynomial functions of ε: $\gamma(\varepsilon) = \Theta(\varepsilon^{t_1})$ and $\sigma^2(\varepsilon) = \Theta(\varepsilon^{t_2})$ for some constants $t_1 \geq 0$ and $t_2 \geq 0$. AGEM and AGEV state that the sample paths that contribute to the highest-order terms in these polynomial functions are not rare.

Definition 6 *(AGEM and AGEV) The estimator $Y(\varepsilon)$ has the AGEM property if $yp(\varepsilon, y) = \Theta(\epsilon^{t_1})$ implies that $p(\varepsilon, y) = \Theta(1)$ (or equivalently, that $y = \Theta(\epsilon^{t_1})$). It has the AGEV property if $[y - \gamma(\varepsilon)]^2 p(\varepsilon, y) = \Theta(\epsilon^{t_2})$ implies that $p(\varepsilon, y) = \Theta(1)$ (or equivalently, that $[y - \gamma(\varepsilon)]^2 = \Theta(\epsilon^{t_2})$).*

5

These properties means that for the realizations y of Y that provide the leading contributions to the estimator, the contributions decrease only because of decreasing values of y, and not because of decreasing probabilities. In a setting where IS is applied and Y is the product of an indicator function by a likelihood ratio (this will be the case in Sections 4.2 and 5), this means that the value of the likelihood ratio when $yp(\varepsilon, y)$ contributes to the leading term must converge at the same rate at this leading term when $\varepsilon \to 0$.

3   Importance Sampling

The aim of IS is to reduce the variance by simulating the model with different probability laws for its input random variables and correcting the estimator by a multiplicative weight called the likelihood ratio to recover an unbiased estimator. In rare-event simulation, the probability laws are changed so that the rare events of interest occur more frequently under the new probability measure. We briefly recall the basic definition of IS; for comprehensive overviews see, e.g., [3, 11, 12].

In a general measure theoretic setting, IS is based on the application of the Radon-Nikodym theorem, and the likelihood ratio corresponds to the Radon-Nikodym derivative. All applications of IS are special cases of this setting.

Consider two probability measures $\mathbb{P}$ and $\mathbb{P}^*$ on a measurable space $(\Omega, \mathcal{A})$, where $\mathbb{P}$ is absolutely continuous with respect to $\mathbb{P}^*$, which means that for all $A \in \mathcal{A}$, $\mathbb{P}^*\{A\} = 0 \Rightarrow \mathbb{P}\{A\} = 0$. Then, the Radon-Nikodym theorem guarantees that for $\mathbb{P}$-almost all $\omega \in \Omega$, the Radon-Nikodym derivative $L(\omega) = (d\mathbb{P}/d\mathbb{P}^*)(\omega)$ exists, and that

$$\mathbb{P}\{A\} = \int_A L(\omega)dP^*(\omega) \qquad \text{for all } A \in \mathcal{A}.$$

In the context of IS, $\mathbb{P}^*$ is called the IS measure and we refer to the random variable $L = L(\omega)$ as the *likelihood ratio*. If $Y = Y(\omega)$ is a random variable defined on $(\Omega, \mathcal{A})$, and if $d\mathbb{P}^*(\omega) > 0$ whenever $Y(\omega)d\mathbb{P}(\omega) > 0$, then

$$E_{\mathbb{P}}[Y] = \int Y(\omega)d\mathbb{P}(\omega) = \int Y(\omega)L(\omega)d\mathbb{P}^*(\omega) = \mathbb{E}_{\mathbb{P}^*}[Y\,L].$$

As a special case, consider a discrete time Markov chain $\{X_j,\ j \geq 0\}$ with a discrete state space $\mathcal{S}$, initial distribution μ over $\mathcal{S}$, and probability transition matrix $\mathbf{P}$. This defines a probability measure over the sample paths of the chain. We are interested in a random variable $Y = g(X_0, X_1, \ldots, X_\tau)$ where τ is a random stopping time and g is a real-valued function. Let μ^* be another initial distribution and let $\mathbf{P}^*$ be another probability transition matrix such that $\mu^*(x_0) \prod_{j=1}^{\tau} \mathbf{P}^*(x_{j-1}, x_j) > 0$ whenever $g(x_0, x_1, \ldots, x_\tau)\mu(x_0) \prod_{j=1}^{\tau} \mathbf{P}(x_{j-1}, x_j) > 0$. Let $\mathbb{P}^*$ be the corresponding probability measure on the Markov chain trajectories. When the sample path is generated from $\mathbb{P}^*$, the likelihood ratio that corresponds to a change from $(\mu, \mathbf{P})$ to $(\mu^*, \mathbf{P}^*)$ and realization $(X_0, \ldots, X_\tau)$ is the random variable

$$L(\omega) = L(X_0, X_1, \ldots, X_\tau) = \frac{\mu(X_0) \prod_{j=1}^{\tau} \mathbf{P}(X_{j-1}, X_j)}{\mu^*(X_0) \prod_{j=1}^{\tau} \mathbf{P}^*(X_{j-1}, X_j)}$$

if $\mu^*(X_0) \prod_{j=1}^{\tau} \mathbf{P}^*(X_{j-1}, X_j) \neq 0$, and 0 otherwise. Hence,

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}}[g(X_0, \ldots, X_\tau)] \\
={}& \sum_{n=0}^{\infty} \sum_{(x_0, x_1, \ldots, x_n) \in \mathcal{S}^n} 1_{\{\tau=n\}} g(x_0, x_1, \ldots, x_n)\mu(x_0) \prod_{j=1}^{n} \mathbf{P}(x_{j-1}, x_j) \\
={}& \sum_{n=0}^{\infty} \sum_{(x_0, x_1, \ldots, x_n) \in \mathcal{S}^n} 1_{\{\tau=n\}} g(x_0, x_1, \ldots, x_n)L(x_0, x_1, \ldots, x_n)\mu^*(x_0) \prod_{j=1}^{n} \mathbf{P}^*(x_{j-1}, x_j) \\
={}& \mathbb{E}_{\mathbb{P}^*}[g(X_0, \ldots, X_\tau)L(X_0, \ldots, X_\tau)].
\end{aligned}$$

Note that $\mathbb{P}^*(X_0, \ldots, X_\tau) = 0$ is required only if $g(X_0, \ldots, X_\tau)\mathbb{P}(X_0, \ldots, X_\tau) = 0$. An IS estimator generates a sample path $X_0, \ldots, X_\tau$ using $\mathbb{P}^*$ and computes

$$Y = g(X_0, X_1, \ldots, X_\tau)L(X_0, X_1, \ldots, X_\tau) \tag{8}$$

as an estimator of $\mathbb{E}_{\mathbb{P}}[Y] = \mathbb{E}_{\mathbb{P}}[g(X_0, X_1, \ldots, X_\tau)]$.

4 Importance Sampling for a Highly Reliable Markovian System

4.1 The Model

We consider an HRMS with c types of components and n_i components of type i, for $i = 1, \ldots, c$. Each component is either in a failed state or an operational state. The *state of the system* is represented by a vector $x = (x^{(1)}, \ldots, x^{(c)})$, where $x^{(i)}$ is the number of *failed* components of type i. Thus, we have a finite state space $\mathcal{S}$ of cardinality $(n_1 + 1)\cdots(n_c + 1)$. We suppose that $\mathcal{S}$ is partitioned in two subsets $\mathcal{U}$ and $\mathcal{F}$, where $\mathcal{U}$ is a decreasing set (i.e., if $x \in \mathcal{U}$ and $x \geq y \in \mathcal{S}$, then $y \in \mathcal{U}$) that contains the state $\mathbf{1} = (0, \ldots, 0)$ in which all the components are operational. We say that $y \prec x$ when $y \leq x$ and $y \neq x$.

We assume that the times to failure and times to repair of the individual components are independent exponential random variables with respective rates

$$\lambda_i(x) = a_i(x)\varepsilon^{b_i(x)} = O(\varepsilon) \quad \text{and} \quad \mu_i(x) = \Theta(1)$$

for type-i components when the current state is x, where $a_i(x)$ is a strictly positive real number and $b_i(x)$ a strictly positive integer for each i. The parameter $\varepsilon \ll 1$ represents the rarity of failures; the failure rates tend to zero when $\varepsilon \to 0$. Failure propagation is allowed: from state x, there is a probability $p_i(x, y)$ (which may depend on ε) that the failure of a type-i component directly drives the system to state y, in which there could be additional component failures. Thus, the net jump rate from x to y is

$$\lambda(x, y) = \sum_{i=1}^{c} \lambda_i(x)p_i(x, y) = O(\varepsilon).$$

Similarly, the repair rate from state x to state y is $\mu(x, y)$ (with possible grouped repairs), where $\mu(x, y)$ does not depend on ε (i.e., repairs are not rare events when they are possible). The system starts in state $\mathbf{1}$ and we want to estimate the probability $\gamma(\varepsilon)$ that it reaches the set $\mathcal{F}$ before returning to state $\mathbf{1}$. Estimating this probability is relevant in many practical situations [11, 12].

This model evolves as a continuous-time Markov chain (CTMC) $(Y(t), t \geq 0\}$, where $Y(t)$ is the system's state at time t. Its canonically embedded discrete time Markov chain (DTMC) is $\{X_j, j \geq 0\}$, defined by $X_j = Y(\xi_j)$ for $j = 0, 1, 2, \ldots$, where $\xi_0 = 0$ and $0 < \xi_1 < \xi_2 < \cdots$ are the jump times of the CTMC. Since the quantity of interest here, $\gamma(\varepsilon)$, does not depend on the jump times of the CTMC, it suffices to simulate the DTMC. This chain $\{X_j, j \geq 0\}$ has transition probability matrix $\mathbf{P}$ with elements

$$\mathbf{P}(x, y) = \mathbb{P}[X_j = y \mid X_{j-1} = x] = \lambda(x, y)/q(x)$$

if the transition from x to y corresponds to a failure and

$$\mathbf{P}(x, y) = \mu(x, y)/q(x)$$

if it corresponds to a repair, where

$$q(x) = \sum_{y \in \mathcal{S}} (\lambda(x, y) + \mu(x, y))$$

is the total jump rate out of x, for all x, y in $\mathcal{S}$. We will use $\mathbb{P}$ to denote the corresponding measure on the sample paths of the DTMC.

7

Let Γ denote the set of pairs $(x, y) \in \mathcal{S}^2$ for which $\mathbf{P}(x, y) > 0$. Our final assumptions are that the DTMC is irreducible on $\mathcal{S}$ and that for every state $x \in \mathcal{S}$, $x \neq \mathbf{1}$, there exists a state $y \prec x$ such that $(x, y) \in \Gamma$ (that is, at least one repairman is active whenever a component is failed).

Again, our goal is to estimate $\gamma(\varepsilon) = \mathbb{P}[\tau_{\mathcal{F}} < \tau_{\mathbf{1}}]$, where $\tau_{\mathcal{F}} = \inf\{j > 0 : X_j \in \mathcal{F}\}$ and $\tau_{\mathbf{1}} = \inf\{j > 0 : X_j = \mathbf{1}\}$. It has been shown [17] that for this model, there is an integer $r > 0$ such that $\gamma(\varepsilon) = \Theta(\varepsilon^r)$, i.e., the probability of interest decreases at a polynomial rate when $\varepsilon \to 0$.

4.2   IS for the HRMS Model

Naive Monte Carlo estimates $\gamma(\varepsilon)$ by simulating samples paths with the transition probability matrix $\mathbb{P}$ and counting the fraction of those paths for which $\tau_{\mathcal{F}} < \tau_{\mathbf{1}}$. But since $\gamma(\varepsilon) = \Theta(\varepsilon^r)$, the relative error of this estimator increases toward infinity when $\varepsilon \to 0$ and something else must be done to obtain a viable estimator.

Several IS schemes have been proposed in the literature for this HRMS model; see, e.g., [4, 15, 17]. Here we limit ourselves to the so-called *simple failure biasing* (SFB), also named *Bias1*. SFB changes the matrix $\mathbf{P}$ to a new matrix $\mathbf{P}^*$ defined as follows. For states $x \in \mathcal{F} \cup \{\mathbf{1}\}$, we have $\mathbf{P}^*(x, y) = \mathbf{P}(x, y)$ for all $y \in \mathcal{S}$, i.e., the transition probabilities are unchanged. For any other state x, a fixed probability ρ is assigned to the set of all failure transitions, and a probability $1 - \rho$ is assigned to the set of all repair transitions. In each of these two subsets, the individual probabilities are taken proportionally to the original ones. Under certain additional assumptions, this change of measure increases the probability of failure when the system is up, in a way that failure transitions are no longer rare events, i.e., $\mathbb{P}^*[\tau_{\mathcal{F}} < \tau_{\mathbf{1}}] = \Theta(1)$.

For a given sample path ending at step $\tau = \min(\tau_{\mathcal{F}}, \tau_{\mathbf{1}})$, the likelihood ratio for this change of measure can be written as

$$L = L(X_0, \ldots, X_\tau) = \frac{\mathbb{P}[(X_0, \ldots, X_\tau)]}{\mathbb{P}^*[(X_0, \ldots, X_\tau)]} = \prod_{j=1}^{\tau} \frac{\mathbf{P}(X_{j-1}, X_j)}{\mathbf{P}^*(X_{j-1}, X_j)}$$

and the corresponding (unbiased) IS estimator of $\gamma(\varepsilon)$ is given by (8), with $g(X_0, \ldots, X_\tau) = 1_{\{\tau_{\mathcal{F}} < \tau_{\mathbf{1}}\}}$. Thus, the random variable $Y(\varepsilon)$ of Section 2 is

$$Y(\varepsilon) = 1_{\{\tau_{\mathcal{F}} < \tau_{\mathbf{1}}\}} L(X_0, \ldots, X_\tau). \tag{9}$$

We will now examine the robustness properties of this estimator under the SFB sampling.

5   Asymptotic Robustness Properties for the HRMS Model Under IS

A characterization of the IS schemes for the HRMS model that satisfy the BRE property was obtained in [15]. AO is weaker than BRE in general. However, our first result states that for the HRMS model, the two are equivalent. This was mentioned without proof in [11].

Theorem 2 *In our HRMS framework, with SFB, AO and BREff are equivalent to BRE.*

Proof. Recall that $\gamma(\varepsilon) = \Theta(\varepsilon^r)$ for some integer $r \geq 0$. It has also been shown in [19] that for this model, $\mathbb{E}[Y^2(\varepsilon)] = \Theta(\varepsilon^s)$ for some $s \leq 2r$, where $Y(\varepsilon)$ is defined in (9). Also $t(\varepsilon) = \Theta(1)$ for static changes of measure such as SFB. The equivalence between AO and BRE then follows from Example 2, and the equivalence between BREff and BRE follows from Example 3. $\square$

A characterization of IS measures that satisfy BNA for the HRMS model is given in [19, 20] and the following relationships between measures of robustness was proved in [20]:

Theorem 3 *In our HRMS framework, BNA implies AGEV, which implies BRE, which implies AGEM. For each of these implications, the converse is not true.*

Our next results characterize the BREEV and AOEV in the HRMS framework. They require additional notation. We will restrict our change of measure for IS to a class $\mathcal{I}$ of measures $\mathbb{P}^*$ defined by a transition probability matrix $\mathbf{P}^*$ with the following properties: whenever $(x, y) \in \Gamma$ and $\mathbf{P}(x, y) = \Theta(\varepsilon^d)$, if $y \succ x \neq \mathbf{1}$, then $\mathbf{P}^*(x, y) = \underline{O}(\varepsilon^{d-1})$, whereas if $x \succ y$ or if $y \succ x = \mathbf{1}$, then $\mathbf{P}^*(x, y) = \underline{O}(\varepsilon^d)$. This class $\mathcal{I}$ was introduced in [19]; these measures increase the probability of each failure transition from a state $x \neq \mathbf{1}$ and satisfy the assumptions of Lemma 1 of [19] (we will use this lemma to prove our next results). From now on, we assume that $\mathbf{P}^*$ satisfies these properties.

We define the following sets of sample paths:

$$
\begin{aligned}
\Delta_m &= \{(x_0, \cdots, x_n) : n \geq 1, x_0 = \mathbf{1}, x_n \in \mathcal{F}, x_j \notin \{\mathbf{1}, F\} \text{ and} (x_{j-1}, x_j) \in \Gamma \text{ for } 1 \leq j \leq n, \\
&\qquad \text{and } \mathbb{P}\{(X_0, \cdots, X_\tau) = (x_0, \cdots, x_n)\} = \Theta(\varepsilon^m)\}; \\
\Delta_{m,k} &= \{(x_0, \cdots, x_n) \in \Delta_m : \mathbb{P}^*\{(X_0, \cdots, X_\tau) = (x_0, \cdots, x_n)\} = \Theta(\varepsilon^k)\}; \\
\Delta'_t &= \bigcup_{\{m,k \,:\, m-k=t\}} \Delta_{m,k};
\end{aligned}
$$

and let s be the integer such that $\sigma^2_{\mathbb{P}^*}(\varepsilon) = \Theta(\varepsilon^s)$.

A necessary and sufficient condition on $\mathbb{P}^*$ for BREEV is as follows. This result means that a path cannot be too rare under the IS measure P^* to verify BREEV. Similar results were obtained under the same conditions for BRE in [15], where it is shown that $k \leq 2m - r$ is needed when $\Delta_{m,k} \neq \emptyset$, and for BNA in [19, 20], where the necessary and sufficient condition is $k \leq 3m/2 - 3s/4$.

Theorem 4 *For an IS measure $\mathbb{P}^* \in \mathcal{I}$, we have BREEV if and only if for all integers k and m such that $m - k < r$ and all $(x_0, \cdots, x_n) \in \Delta_{m,k}$,*

$$
P^*\{(X_0, \cdots, X_\tau) = (x_0, \cdots, x_n)\} = \underline{O}(\varepsilon^{4m/3 - 2s/3}).
$$

In other words, we must have $k \leq 4m/3 - 2s/3$ whenever $\Delta_{m,k} \neq \emptyset$.

Proof. (a) Necessary condition. Suppose that there exist $k, m \in \mathbb{N}$ and $(x_0, \cdots, x_n) \in \Delta_{m,k}$ such that $k = 4m/3 - 2s/3 + k'$ with $k' > 0$ and $m - k < r$. This means that $\mathbb{P}^*\{(X_0, \cdots, X_\tau) = (x_0, \cdots, x_n)\} = \Theta(\varepsilon^{4m/3 - 2s/3 + k'})$. Then we have

$$
\begin{aligned}
\mathbb{E}[(Y(\varepsilon) - \gamma(\varepsilon))^4] &\geq [L(x_0, \cdots, x_n) - \gamma(\varepsilon)]^4 \, \mathbb{P}^*\{(X_0, \cdots, X_\tau) = (x_0, \cdots, x_n)\} \\
&= \Theta(\varepsilon^{4(m-k)+k}) \\
&= \Theta(\varepsilon^{2s - 3k'}).
\end{aligned}
$$

Thus $\mathbb{E}[(Y(\varepsilon) - \gamma(\varepsilon))^4]/\sigma^4_{\mathbb{P}^*}(\varepsilon) = \underline{O}(\varepsilon^{-2k'})$, which is unbounded when $\varepsilon \to 0$.

(b) Sufficient condition. Let $(x_0, \cdots, x_n) \in \Delta'_t$ such that $t < r$ (i.e., $m - k < r$). Since

$$
\mathbb{P}^*\{(X_0, \cdots, X_\tau) = (x_0, \cdots, x_n)\} = \underline{O}(\varepsilon^{4m/3 - 2s/3})
$$

for all $(x_0, \cdots, x_n) \in \Delta_{m,k}$ if $m - k < r$, we have

$$
(L(x_0, \cdots, x_n) - \gamma(\varepsilon))^4 \mathbb{P}^*\{(X_0, \cdots, X_\tau) = (x_0, \cdots, x_n)\} = \frac{\Theta(\varepsilon^{4m})}{\Theta(\varepsilon^{4k})}\Theta(\varepsilon^k) = O(\varepsilon^{2s}).
$$

Using the fact that $\sum_{t<r} |\Delta'_t| < \infty$ and the first part of Lemma 1 of [19], we have

$$
\sum_{t<r} \sum_{(x_0, \cdots, x_n) \in \Delta'_t} (L(x_0, \cdots, x_n))^4 P^*\{(X_0, \cdots, X_{\tau_F}) = (x_0, \cdots, x_n)\} = O(\varepsilon^{2s}).
$$

Since

$$
\sum_{t=r}^{\infty} \sum_{(x_0, \cdots, x_n) \in \Delta'_t} \mathbb{P}^*\{(X_0, \cdots, X_\tau) = (x_0, \cdots, x_n)\} \leq 1
$$

and $\displaystyle\sum_{(x_0,\cdots,x_n)\in\Delta'_t}\mathbb{P}^*\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}\leq 1$ for all t, Lemma 1 of [19] implies that

$$\sum_{t=r}^{\infty}\sum_{(x_0,\cdots,x_n)\in\Delta'_t}(L(x_0,\cdots,x_n)-\gamma(\varepsilon))^4\mathbb{P}^*\{(X_0,\cdots,X_\tau)=(x_0,\cdots,x_n)\}$$

$$\leq\sum_{t=r}^{\infty}\sum_{(x_0,\cdots,x_n)\in\Delta'_t}(\gamma(\varepsilon)^4+4\gamma(\varepsilon)^3\kappa\eta^t\varepsilon^t+6\gamma^2(\varepsilon)\kappa^2\eta^{2t}\varepsilon^{2t}+4\gamma(\varepsilon)\kappa^3\eta^{3t}\varepsilon^{3t}+\kappa^4\eta^{4t}\varepsilon^{4t})\cdot$$
$$\mathbb{P}^*\{(X_0,\cdots,X_\tau)=(x_0,\cdots,x_n)\}$$

$$=\gamma^4(\varepsilon)\sum_{t=r}^{\infty}\sum_{(x_0,\cdots,x_n)\in\Delta'_t}\mathbb{P}^*\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}$$
$$+\sum_{t=r}^{\infty}(4\gamma^3(\varepsilon)\kappa\eta^t\varepsilon^t+6\gamma^2(\varepsilon)\kappa^2\eta^{2t}\varepsilon^{2t}+4\gamma(\varepsilon)\kappa^3\eta^{3t}\varepsilon^{3t}+\kappa^4\eta^{4t}\varepsilon^{4t})\cdot$$
$$\sum_{(x_0,\cdots,x_n)\in\Delta'_t}\mathbb{P}^*\{(X_0,\cdots,X_\tau)=(x_0,\cdots,x_n)\}$$

$$\leq\gamma^4+4\gamma^3(\varepsilon)\kappa\sum_{t=r}^{\infty}(\eta\varepsilon)^t+6\gamma^2(\varepsilon)\kappa^2\sum_{t=r}^{\infty}(\eta^2\varepsilon^2)^t4\gamma(\varepsilon)\kappa^3\sum_{t=r}^{\infty}(\eta^3\varepsilon^3)^t+\kappa^4\sum_{t=r}^{\infty}(\eta^4\varepsilon^4)^t$$

$$=\Theta(\varepsilon^{4r})+\Theta(\varepsilon^{4r})+\Theta(\varepsilon^{4r})+\Theta(\varepsilon^{4r})+\Theta(\varepsilon^{4r})$$

$$=O(\varepsilon^{2s}),$$

because $2r\geq s$. $\qquad\square$

Theorem 5 *BREEV, BREffEV, and AOEV are equivalent.*

Proof. This follows again directly from Example 2, using the fact that $\sigma^2(\varepsilon)=\Theta(\varepsilon^s)$ and $\mathbb{E}[S_n^4(\varepsilon)]=\Theta(\varepsilon^t)$ with $t\leq 2s$, and from Example 3 since $t(\varepsilon)=\Theta(1)$. $\qquad\square$

Next we show that BREEV and AOEV are the strongest properties in our list.

Theorem 6 *BREEV implies BNA.*

Proof. This is a direct consequence of the necessary and sufficient conditions over the paths for the BNA and BREEV properties. These conditions are that for all k and m such that $m-k<r$, whenever $\Delta_{m,k}$ is non-empty, we must have $k\leq 4m/3-2s/3$ for BREEV and $k\leq 3m/2-3s/4$ for BNA. But $4m/3-2s/3=8/9(3m/2-3s/4)$, so the theorem is proved if we always have $3m/2-3s/4\geq 0$, i.e., $2m\geq s$, which is true since $2m\geq 2r\geq s$. $\qquad\square$

The following counter-example show that the converse is not true: there exist systems and IS measures $\mathbb{P}^*$ for which BNA is verified but not BREEV.

Example 5 *Consider the example of Figure 1, using SFB failure biasing as shown in Figure 2. The states where the system is down are colored in grey.*

For this model, as it can be easily seen in Figure 1, $r=6$ and Δ_6 is comprised of the single path $(<2,2>,<1,2>,<0,2>)$. Moreover, $s=12$ and the sole path in Δ such that

$$\frac{\mathbb{P}^2\{(X_0,\cdots,X_\tau)=(x_0,\cdots,x_n)\}}{\mathbb{P}^*\{(X_0,\cdots,X_\tau)=(x_0,\cdots,x_n)\}}=\Theta(\varepsilon^{12})$$

is the path in Δ_6 for which Figure 2 shows that it is $\Theta(1)$ under probability measure $\mathbb{P}^*$. It can also be readily checked that $k\leq 3m/2-3s/4$ for all paths, meaning that BNA is verified.

However, the path $(<2,2>,<2,1>,<2,0>)$ is in $\Delta_{m,k}$ with $m=14$ and $k=12$. Then $12=k>4m/3-2s/3=32/3$, so the necessary and sufficient condition of Theorem 4 is not verified.
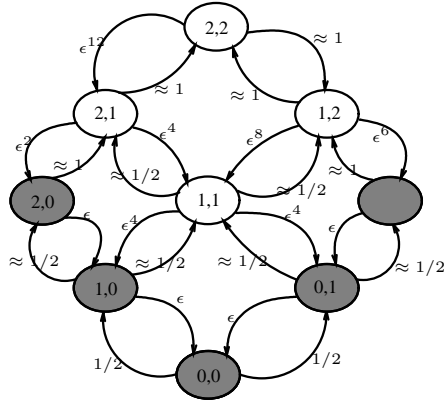
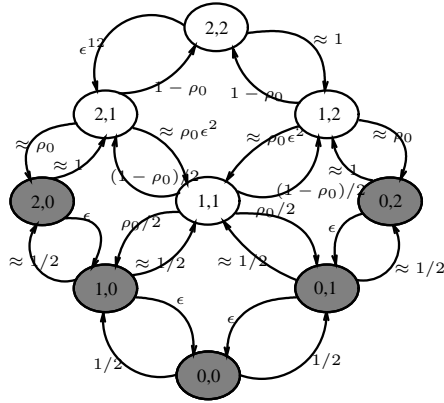Figure 1: A two-dimensional model with its transition probabilities.

Figure 2: A two-dimensional example with SFB transition probabilities.

6 Conclusions

We have extended the hiecharchy of robustness properties of IS estimators for an HRMS model by adding AO and the newly defined BREEV, BREffEV, and AOEV, that assert the stability of the relative size of the confidence interval for independent samples, as rarity increases. The complete hierarchy can be summarized as:

$$(\text{BREEV} \Leftrightarrow \text{BREffEV} \Leftrightarrow \text{AOEV}) \Rightarrow \text{BNA} \Rightarrow \text{AGEV} \Rightarrow (\text{BRE} \Leftrightarrow \text{BREff} \Leftrightarrow \text{AO}) \Rightarrow \text{AGEM}.$$

All these properties have some practical relevance and understanding the links between them is certainly of high interest. BREEV is the strongest, but it may be difficult to verify in some applications. A direction of future research is to study this hierarchy of properties in more general (or different) settings; for example in situations where $\gamma(\varepsilon)$ converges to zero exponentially fast. It is already known that AO is generally not equivalent to BRE in this case. What about the other implications in the hierarchy?

References

[1] S. Asmussen. Large deviations in rare events simulation: Examples, counterexamples, and alternatives. In K.-T. Fang, F. J. Hickernell, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo*

Methods 2000, pages 1–9, Berlin, 2002. Springer-Verlag.

[2] V. Bentkus and F. Götze. The Berry-Esseen bound for Student's statistic. *The Annals of Probability*, 24(1):491–503, 1996.

[3] J. A. Bucklew. *Introduction to Rare Event Simulation*. Springer-Verlag, New York, 2004.

[4] H. Cancela, G. Rubino, and B. Tuffin. MTTF estimation by Monte Carlo methods using Markov models. *Monte Carlo Methods and Applications*, 8(4):312–341, 2002.

[5] H. Cancela, G. Rubino, and B. Tuffin. New measures of robustness in rare event simulation. In M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, editors, *Proceedings of the 2005 Winter Simulation Conference*, pages 519–527, 2005.

[6] W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. 2*. Wiley, New York, second edition, 1971.

[7] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, New York, 2004.

[8] P. Glasserman, P. Heidelberger, P. Shahabuddin, and T. Zajic. A large deviations perspective on the efficiency of multilevel splitting. *IEEE Transactions on Automatic Control*, AC-43(12):1666–1679, 1998.

[9] P. W. Glynn and D. L. Iglehart. Importance sampling for stochastic simulations. *Management Science*, 35:1367–1392, 1989.

[10] A. Goyal, P. Shahabuddin, P. Heidelberger, V. F. Nicola, and P. W. Glynn. A unified framework for simulating Markovian models of highly reliable systems. *IEEE Transactions on Computers*, C-41:36–51, 1992.

[11] P. Heidelberger. Fast simulation of rare events in queueing and reliability models. *ACM Transactions on Modeling and Computer Simulation*, 5(1):43–85, 1995.

[12] S. Juneja and P. Shahabuddin. Rare event simulation techniques: An introduction and recent advances. In S. G. Henderson and B. L. Nelson, editors, *Stochastic Simulation*, Handbooks of Operations Research and Management Science. Elsevier Science, circa 2006. chapter 11, to appear.

[13] P. L'Ecuyer, V. Demers, and B. Tuffin. Rare-events, splitting, and quasi-Monte Carlo. *ACM Transactions on Modeling and Computer Simulation*, 2006. submitted.

[14] E. E. Lewis and F. Böhm. Monte Carlo simulation of Markov unreliability models. *Nuclear Engineering and Design*, 77:49–62, 1984.

[15] M. K. Nakayama. General conditions for bounded relative error in simulations of highly reliable Markovian systems. *Advances in Applied Probability*, 28:687–727, 1996.

[16] W. Sandmann. Relative error and asymptotic optimality in estimating rare event probabilities by importance sampling. In *Proceedings of the OR Society Simulation Workshop (SW04) held in cooperation with the ACM SIGSIM, Birmingham, UK, March 23–24, 2004*, pages 49–57. The Operational Research Society, 2004.

[17] P. Shahabuddin. Importance sampling for the simulation of highly reliable Markovian systems. *Management Science*, 40(3):333–352, 1994.

[18] D. Siegmund. Importance sampling in the Monte Carlo study of sequential tests. *The Annals of Statistics*, 4:673–684, 1976.

[19] B. Tuffin. Bounded normal approximation in simulations of highly reliable Markovian systems. *Journal of Applied Probability*, 36(4):974–986, 1999.

[20] B. Tuffin. On numerical problems in simulations of highly reliable Markovian systems. In *Proceedings of the 1st International Conference on Quantitative Evaluation of SysTems (QEST)*, pages 156–164, University of Twente, Enschede, The Netherlands, September 2004. IEEE CS Press.

[21] P. van Beek. An application of Fourier methods to the problem of sharpening the Berry-Esseen inequality. *Zur Wahrscheinlichkeitstheorie verwandte Gebiete*, 23:187–196, 1972.

[22] M. Villén-Altamirano and J. Villén-Altamirano. On the efficiency of RESTART for multidimensional systems. *ACM Transactions on Modeling and Computer Simulation*, 16(3):251–279, 2006.