Bounded Normal Approximation in Simulations of Highly Reliable Markovian Systems

Bruno Tuffin
IRISA
Campus de Beaulieu
35042 Rennes Cedex
France
e-mail: btuffin@irisa.fr

Abstract

In this paper, we give necessary and sufficient conditions that ensure the validity of confidence intervals, based on the central limit Theorem, in simulations of highly reliable Markovian systems. We take recourse to simulation because of the frequent huge state space in practical systems. So far the literature has focused on the property of bounded relative error. In this paper we focus on "bounded normal approximation" which asserts that the approximation of the normal law, suggested by the central limit Theorem, does not deteriorate as the reliability of the system increases. We see that the set of systems with bounded normal approximation is (strictly) included in the set of systems with bounded relative error.

Keywords: Simulation, Normal Approximation, Markov Chains, Highly Reliable Systems.

AMS Subject Classification: Primary 65C05
Secondary 60J10; 60K10

1 Introduction

Fault tolerant multi-components systems (which tolerate also fault propagation), such as computer or telecommunication systems, are becoming more and more reliable. Such systems are often represented by Markovian models. Direct computational time of dependability metrics in these models, like the MTTF (Mean Time

To Failure) or the availability, is very expensive. Also, approximate numerical techniques like state lumping and unlumping e.g. [4] and state aggregation and bounding e.g. [10] require considerable computer time and memory. Crude Monte Carlo simulation (i.e. sampling from the embedded Markov chain using the original probability measure) is inefficient because of the rarity of the system failure events. Thus we use variance reduction methods, principally importance sampling techniques. A general description of this technique can be found in [8]. In the literature, a number of schemes have been proposed for highly reliable Markovian systems. These include Bias1 failure biasing [9] (also called simple failure biasing), balanced failure biasing [13], Bias2 failure biasing [5] and failure distance biasing [2]. All these schemes increase the probability of transitions corresponding to component failures and thus make that the system failure events happen more often. The estimates are then adjusted to make them unbiased. Shahabuddin [13] introduced the notion of *bounded relative error*. If an estimator enjoys this property, we only need a fixed number of replications to obtain a confidence interval having a fixed relative error, no matter how rarely system failures occur. Shahabuddin [13] showed that balanced failure biasing has the bounded relative error property for a large class of reliability models; the paper also showed that the Bias1 failure biasing scheme has the bounded relative error property for a special class of systems called balanced systems. A more general study of when or when not a scheme has bounded relative error may be found in [12].

The confidence interval is developed from the approximation of the distribution of a normalized sum of independent and identically distributed random variables by the normal distribution, as suggested by the central limit Theorem. In this paper, we show that we also have to consider the validity of this normal approximation. We develop conditions under which the error in the normal approximation remains bounded as systems failures become rarer - we call the latter *bounded normal approximation*. To obtain bounded normal approximation is as important as to obtain bounded relative error, because it justifies the validity of the confidence interval, and hence the use of the method, for a fixed number of observations, as the reliability increases. Moreover, we prove that bounded normal approximation implies bounded relative error.

This paper is organized as follows: in section 2, we recall the model specification given by Nakayama in [12] which is a slightly modified version of the one originally presented by Shahabuddin in [13]. In section 3 we give the theorem that gives conditions under which one gets bounded normal approximation. Section 4 shows that schemes with bounded normal approximation are also ones with bounded relative error and that the inclusion is strict. It is also shown in the same section that the balanced failure biasing scheme has the bounded normal approximation property for a large class of reliability models and that the three other commonly used schemes have the property for a more restrictive class of models. Finally we conclude in section 5.

2 Model presentation

We use the notations of [13] and [12]. A function f is said to be $o(\varepsilon^d)$ if $f(\varepsilon)/\varepsilon^d \to 0$ as $\varepsilon \to 0$. Similarly, $f(\varepsilon) = O(\varepsilon^d)$ if $|f(\varepsilon)| \leq c_1\varepsilon^d$ for some constant $c_1 > 0$ for all ε sufficiently small. It is said to be $\underline{O}(\varepsilon^d)$ if $|f(\varepsilon)| \geq c_2\varepsilon^d$ for some constant $c_2 > 0$ for all ε sufficiently small. Finally, $f(\varepsilon) = \Theta(\varepsilon^d)$ if $f(\varepsilon) = \underline{O}(\varepsilon^d)$ and $f(\varepsilon) = O(\varepsilon^d)$.

We suppose that the system has C types of components, with n_i components of type i. The total number of components is then $N = \sum_{i=1}^{C} n_i$. The system is subject to random failures and repairs with exponential laws. The model is given by a continuous time Markov chain (CTMC) $(Y_t)_{t\geq 0}$ defined on the finite state space S where each $x \in S$ gives for all $i = 1, \ldots, C$, the number of operational components (also called up components) of type i, $n_i(x)$. We label the state with all components up as $\mathbf{1}$. We suppose that this is the initial state. S is partitioned into two subsets U and F, where U denotes the set of up states and F the set of down states. We assume that if $x \in U$ and $y \in S$ with $n_i(y) \geq n_i(x)$ for all component types i, then $y \in U$. Of course, $\mathbf{1} \in U$. Failure propagations are allowed. Let $p(y; x, i)$ be the probability that, if the system is in state x and a component of type i fails, the system goes directly to state y by means of propagations. A transition (x, y) from a state x to a state y is said to be a failure transition if $\forall 1 \leq i \leq C$, $n_i(y) \leq n_i(x)$, with a strict inequality for some type i; this is denoted $y \succ x$. We define in an analogous way the repair transitions, which we denote as $y \prec x$. Let Γ be the set of possible transitions. When we are in state x, a repair occurs to state y with rate $\mu(x, y)$. A failure of a component of type i occurs exponentially in state x with rate $\lambda_i(x)$. Given that failures are rare, we introduce a rarity parameter $\varepsilon > 0$, such that $\varepsilon \ll 1$ such that

$$\lambda_i(x) = a_i(x)\varepsilon^{b_i(x)},$$

where $a_i(x) \geq 0$ and $b_i(x) \geq 1$ are independent of ε. In the same way we suppose that

$$p(y;x,i) = c_i(x,y)\varepsilon^{d_i(x,y)}$$

where $d_i(x,y) \geq 0$ is integer-valued, $c_i(x,y) \geq 0$ and $\sum_{y\in S} p(y;x,i) = 1$. We assume that repair rates $\mu(x,y)$ are independent of ε.

The infinitesimal generator of Y, denoted by $Q = (q(x,y))_{x,y\in S}$, is given by

$$q(x,y) = \begin{cases} \sum_{k=1}^{C} n_k(x)\lambda_k(x)p(y;x,k) & \text{if } y \succ x \\ \mu(x,y) & \text{if } y \prec x \\ 0 & \text{elsewhere} \end{cases}$$

for $x \neq y$, and $-q(x,x) = \sum_{x\neq y} q(x,y)$. Define $q(x) = -q(x,x)$ and let us denote by X the canonically embedded discrete time Markov chain (DTMC) and by $\boldsymbol{P}$ its transition matrix. If we call $b_0 = \min_{1\leq i\leq C} b_i(\mathbf{1})$ and

$$b(x,y) = \begin{cases} \min_{i=1}^{C}\{b_i(x) + d_i(x,y) : n_i(x)a_i(x)p(y;x,i) > 0\} & \text{if } y \succ x \\ 0 & \text{if } y \prec x \end{cases}$$

is the exponent of the order of magnitude of the rate of the transition (x,y), we have [12] that for any $(x,y) \in \Gamma$,

$$\boldsymbol{P}(x,y) = \begin{cases} \Theta(\varepsilon^{b(x,y)}) & \text{if } x \neq \mathbf{1} \\ \Theta(\varepsilon^{b(x,y)-b_0}) & \text{if } x = \mathbf{1}. \end{cases}$$

Define Φ as the corresponding measure on the sample paths of the DTMC. It is known that the MTTF (Mean Time To Failure) can be expressed by the ratio [5]

$$MTTF = \frac{E_\Phi\left(\sum_{k=0}^{\min(\tau_F,\tau_{\mathbf{1}})-1} 1/q(X_k)\right)}{E_\Phi(1_{[\tau_F<\tau_{\mathbf{1}}]})}, \tag{1}$$

where τ_F is the hitting time of the DTMC X to set F and $\tau_{\mathbf{1}}$ the hitting time to state $\mathbf{1}$. This performance measure is estimated by means of regenerative simulation. As the numerator can be efficiently estimated with crude Monte Carlo simulation, most papers in this area focus on the estimation of $\gamma = E_\Phi(1_{[\tau_F<\tau_{\mathbf{1}}]})$. Importance sampling is used in [5], [13], [12]. As a matter of fact, a crude Monte Carlo simulation is statistically inefficient, so very large sample sizes are required to achieve accurate estimators of γ as $\varepsilon \to 0$. We choose a new matrix $\boldsymbol{P}'$ and evaluate

$$\gamma = E_{\Phi'}(1_{[\tau_F<\tau_{\mathbf{1}}]}L)$$

with for all path $(x_0, \cdots x_n)$, the likelihood function L is

$$L(x_0, \cdots x_n) = \frac{\Phi\{(X_0, \cdots, X_{\tau_F}) = (x_0, \cdots, x_n)\}}{\Phi'\{(X_0, \cdots, X_{\tau_F}) = (x_0, \cdots, x_n)\}}$$

and Φ' is the measure corresponding to matrix $\boldsymbol{P'}$. The most commonly used choices are balanced failure biasing [13], Bias1 failure biasing [9], Bias2 failure biasing [5] and failure distance biasing [2].

We suppose that the system verifies the three following properties:

- A1: the DTMC X is irreducible on S.
- A2: for every state $x \neq \mathbf{1} \in S$, there exists a state y such that $y \prec x$ and $(x, y) \in \Gamma$.
- A3: for each state $z \in F$, such that $(\mathbf{1}, z) \in \Gamma$, $q(\mathbf{1}, z) = o(\varepsilon^{b_0})$.

In [13], the author proves the following result.

Theorem 1 (Shahabuddin(1994)) *There exists a strictly positive constant r such that $\gamma = \Theta(\varepsilon^r)$.*

In the same paper, the concept of bounded relative error is defined as follows:

Definition 1 (Shahabuddin(1994)) *Define $\sigma^2_{\Phi'}$ as the variance of the random variable $1_{[\tau_F < \tau_{\mathbf{1}}]}L$ under probability measure Φ' (which has mean γ) and z_δ as the $1 - \delta/2$ quantile of the standard normal distribution (i.e., mean 0 and variance 1). Then the relative error for a sample size I is defined by*

$$RE = z_\delta \frac{\sqrt{\sigma^2_{\Phi'}/I}}{\gamma}.$$

We say that we have a bounded relative error if RE remains bounded as $\varepsilon \to 0$.

In [13, Proposition 3], it is shown that any $\boldsymbol{P'}$ with elements that are independent of ε gives bounded relative error. From this, it is shown that balanced failure biasing gives bounded relative error for all systems and that Bias1 failure biasing scheme gives bounded relative error for balanced systems (i.e. systems for which failure transitions from state **1** have order 1 probabilities whereas failures from other states have probabilities of the same order of ε). The fact that failure distance biasing and Bias2 failure biasing schemes give bounded relative error for balanced systems can also be inferred easily from Proposition 3 of [13].

Nakayama [12] gives more general conditions of when or when not a biasing scheme gives bounded relative error. We now review the main result of that paper.

Let us denote by Δ_m the set of paths from $\mathbf{1}$ to F without returning to state $\mathbf{1}$ and with probability $\Theta(\varepsilon^m)$, that is

$$\Delta_m = \{(x_0, \cdots, x_n) : \; n \geq 1, x_0 = \mathbf{1}, x_n \in F, x_i \notin \{\mathbf{1}, F\} \text{ for } 1 \leq i \leq n{-}1, (x_i, x_i{+}1) \in \Gamma$$

$$\text{and } \Phi\{(X_0, \cdots, X_{\tau_F}) = (x_0, \cdots, x_n)\} = \Theta(\varepsilon^m)\}.$$

We can then obtain a necessary and sufficient condition on the importance sampling measure to have a bounded relative error, which basically says that failures must not be excessively rare under Φ':

Theorem 2 (Nakayama(1996)) *Consider any importance sampling measure Φ' corresponding to a transition matrix P' such that for any $(x,y) \in \Gamma$, if $P(x,y) = \Theta(\varepsilon^d)$, then $P'(x,y) = \underline{O}(\varepsilon^d)$. Then we have bounded relative error if and only if for all $(x_0, \cdots, x_n) \in \Delta_m$, $r \leq m \leq 2r-1$,*

$$\Phi'\{(X_0, \cdots X_{\tau_F}) = (x_0, \cdots, x_n)\} = \underline{O}(\varepsilon^{2m-2r}).$$

3 Normal Approximation

We need the following important result concerning the convergence speed of the Student's statistic to the normal law in the central limit theorem. The proof for the case where σ is known can be found in [3].

Theorem 3 (Berry-Esseen) *[1] Let $\rho = E(|X - E(X)|^3)$, $\sigma^2 = E((X - E(X))^2)$ and $\mathcal{N}(x)$ be the standard normal distribution. For $X_1, \cdots, X_I$ i.i.d. copies of X, define $\bar{X}_I = I^{-1}\sum_{i=1}^{I} X_i$, $\hat{\sigma}_I^2 = I^{-1}\sum_{i=1}^{I}(X_i - \bar{X}_I)^2$ and let F_I be the distribution of the centered and normalized sum $(X_1 + \cdots + X_I)/(\hat{\sigma}_I\sqrt{I}) - E(X)\sqrt{I}/\hat{\sigma}_I$. Then there exists an absolute constant $c > 0$ such that, for each x and I*

$$|F_I(x) - \mathcal{N}(x)| \leq \frac{c\rho}{\sigma^3\sqrt{I}}.$$

Definition 2 *If $\rho_{\Phi'} = E_{\Phi'}\left(\left|1_{[\tau_F < \tau_{\mathbf{1}}]}L - E_{\Phi'}(1_{[\tau_F<\tau_{\mathbf{1}}]}L)\right|^3\right)$ denotes the third-order absolute moment and $\sigma_{\Phi'}$ the standard deviation of the random variable $1_{[\tau_F<\tau_{\mathbf{1}}]}L$ under probability measure Φ', then we say that we have bounded normal approximation if $\rho_{\Phi'}/\sigma_{\Phi'}^3$ remains bounded as $\varepsilon \to 0$.*

This property is essential because if it does not hold, the required number of observations necessary to perform a good normal approximation, and then a good confidence interval coverage, may increase with the system reliability (assuming that $\sigma_{\Phi'}$ is at least well approximated). For centered intervals, if $\rho_{\Phi'}/\sigma^3_{\Phi'}$ is bounded, the coverage error is controlled from Berry-Esseen Theorem because, if z_δ is the $1-\delta/2$ quantile of the standard normal distribution,

$$|[F_I(z_\delta) - F_I(-z_\delta)] - (1-\delta)| \leq |F_I(z_\delta) - \mathcal{N}(z_\delta)| + |F_I(-z_\delta) - \mathcal{N}(-z_\delta)| \leq 2\frac{c\rho_{\Phi'}}{\sigma^3_{\Phi'}\sqrt{I}}.$$

Let us now study if bounded $\rho_{\Phi'}/\sigma^3_{\Phi'}$ is also a necessary condition to control the confidence interval coverage. It is known (see [6] or [7]) that, if the observations X_i $(i = 1, \cdots, I)$ in Theorem 3 come from a nonsingular distribution (which is not true in our application), we have the Edgeworth expansion

$$\begin{aligned} F_I(x) - \mathcal{N}(x) = \mathcal{M}(x)\Bigg[&\frac{\mu_3}{6\sigma^3\sqrt{I}}(2x^2+1) + \frac{1}{12I}\left(\frac{\mu_4}{\sigma^4} - 3\right)(-3x + x^3) \\ &+\frac{1}{18I}\left(\frac{\mu_3}{\sigma^3}\right)^2(3x - 2x^3 - x^5) - \frac{1}{4I}(3x + x^3)\Bigg] + o(1/I), \end{aligned}$$

where $\mathcal{M}$ is the density of the standard normal distribution and $\mu_j = E((X - E(X))^j)$ for $j = 3, 4$. A similar type of expansion can be found for the case where σ is known [7]. This *suggests* that the quality of the normal approximation depends on μ_3/σ^3 in the term in $1/\sqrt{I}$ in the Edgeworth expansion of $F_I(x) - \mathcal{N}(x)$, term very similar to the fraction ρ/σ^3 in the Berry-Esseen Theorem. Even for centered confidence intervals with confidence level $1-\delta$, the quality of the coverage may depend on μ_3/σ^3: the difference between the nominal and the actual coverage for nonsingular distributions is

$$\begin{aligned} F_I(z_\delta) - F_I(-z_\delta) - (1-\delta) &= \mathcal{M}(z_\delta)\Bigg[\frac{1}{9I}\left(\frac{\mu_3}{\sigma^3}\right)^2(3z_\delta - 2z_\delta^3 - z_\delta^5) - \frac{1}{2I}(3z_\delta + z_\delta^3) \\ &\quad +\frac{1}{6I}\left(\frac{\mu_4}{\sigma^4} - 3\right)(-3z_\delta + z_\delta^3)\Bigg] + o(1/I), \end{aligned} \tag{2}$$

which depends on the quantity μ_3/σ^3, similar to ρ/σ^3.

To see if bounded $\rho_{\Phi'}/\sigma^3_{\Phi'}$ is also a necessary condition, we then study (abusively) this Edgeworth expansion. Define $\mu_{3,\Phi'}$ and $\mu_{4,\Phi'}$ as the third and fourth-order moments of the random variable $1_{[\tau_F<\tau_{\mathbf{1}}]}L$. We can see from (2) that the coverage error is in most cases uncontrolled when $\rho_{\Phi'}/\sigma^3_{\Phi'}$ is unbounded, as, except in very marginal cancellation cases, $\mu_{3,\Phi'}/\sigma^3_{\Phi'}$ is unbounded if and only if $\rho_{\Phi'}/\sigma^3_{\Phi'}$ is. We also remark that, for centered intervals, bounded $\mu_{3,\Phi'}/\sigma^3_{\Phi'}$ and $\mu_{4,\Phi'}/\sigma^4_{\Phi'}$ do not imply a bounded coverage error as the term $o(1/I)$ depends also on ε and may be unbounded, so only the Berry-Esseen Theorem ensures that the coverage error is limited as $\varepsilon \to 0$.

Let us define now a class of importance sampling measures. This class increases the probability of each failure transition from a state $x \neq \mathbf{1}$.

Definition 3 *Let $\mathcal{I}$ be the class of measures Φ' corresponding to matrix $\boldsymbol{P}'$ defined as follows: for all $(\omega, y) \in \Gamma$, $\omega \neq \mathbf{1}$ and $y \succ \omega$,*

$$\text{if } \boldsymbol{P}(\omega, y) = \Theta(\varepsilon^d), \text{ then } \boldsymbol{P}'(\omega, y) = \underline{O}(\varepsilon^{d-1})$$

and for all (ω, y) with either $y \prec \omega$ or $y \succ \omega$ and $\omega = \mathbf{1}$,

$$\text{if } \boldsymbol{P}(\omega, y) = \Theta(\varepsilon^d), \text{ then } \boldsymbol{P}'(\omega, y) = \underline{O}(\varepsilon^d).$$

This definition is similar to Definition 2 of [12]. The difference is that the new failure probabilities from $\omega \neq \mathbf{1}$ are in $\underline{O}(\varepsilon^{d-1})$ instead of $\underline{O}(\varepsilon^d)$ in [12], to ensure that $|\Delta'_t| < \infty \ \forall t$, where Δ'_t is defined below. Denote

$$\Delta = \bigcup_{m=r}^{\infty} \Delta_m.$$

For Φ' an importance sampling measure, denote

$$\Delta_{m,k} = \{(x_0, \cdots, x_n) \in \Delta \ : \ \Phi\{(X_0, \cdots, X_{\tau_F}) = (x_0, \cdots, x_n)\} = \Theta(\varepsilon^m)$$
$$\text{and } \Phi'\{(X_0, \cdots, X_{\tau_F}) = (x_0, \cdots, x_n)\} = \Theta(\varepsilon^k)\},$$
$$\Delta'_t = \bigcup_{m,k \ : \ m-k=t} \Delta_{m,k},$$

and let s be the integer such that $\sigma^2_{\Phi'} = \Theta(\varepsilon^s)$. In most cases,

$$s = \min \left\{ j \in \mathbb{N} : \exists (x_0, \cdots, x_n) \in \Delta, \ \frac{\Phi^2}{\Phi'}\{(X_0, \cdots, X_{\tau_F}) = (x_0, \cdots, x_n)\} = \Theta(\varepsilon^j) \right\},$$

but there also can be cancellation of the highest order terms of γ and $E_{\Phi'}(1_{[\tau_F < \tau_{\mathbf{1}}]} L^2)$ when these quantities are of the same order of magnitude, so that $s > 2r$ may occur.

A necessary and sufficient condition on Φ' to obtain a bounded normal approximation is the following:

Theorem 4 *The normal approximation is bounded for a fixed number of observations and a measure $\Phi' \in \mathcal{I}$ if and only if $\forall k, m$ such that $m - k < r$, $(x_0, \cdots, x_n) \in \Delta_{m,k}$,*

$$\Phi'\{(X_0, \cdots, X_{\tau_F}) = (x_0, \cdots, x_n)\} = \underline{O}(\varepsilon^{3m/2 - 3s/4})$$

(i.e. $k \leq 3m/2 - 3s/4$), and, in the cancellation case (i.e. $s > 2r$), if we have also the conditions
(i) $\sum_{t=r+1}^{\infty} \sum_{(x_0,\cdots,x_n)\in\Delta'_t} \Phi'\{(X_0,\cdots,X_{\tau_F}) = (x_0,\cdots,x_n)\} = O(\varepsilon^{3s/2-3r})$,
and (ii) $\forall k \geq 0$, *on the subset* $\Delta_{r+k,k}$ *of* Δ'_r, *the cancellation is of order* $l_{r,k}$ *(i.e.* $\sum_{(x_0,\cdots,x_n)\in\Delta_{r+k,k}} |L(x_0,\cdots,x_n) - \gamma| = \Theta(\varepsilon^{l_{r,k}})$*) with* $l_{r,k} \geq s/2 - k/3$.

The additional conditions (i) and (ii) for the cancellation case mean that each path in Δ'_t with $t \geq r$ is sufficiently rare to ensure that the cancellation has no incidence on the bounded normal approximation property. Before proving this theorem, let us demonstrate the following lemma:

Lemma 1 *If* $\Phi' \in \mathcal{I}$, *then*

- $|\Delta'_t| \leq |S|^{tN+1} < +\infty$;
- *For all* ε *sufficiently small,* $\forall t$ *and* $(x_0,\cdots,x_n) \in \Delta'_t$, $L(x_0,\cdots,x_n) \leq \kappa\eta^t\varepsilon^t$, *where* κ *and* η *are strictly positive constants independent of* ε *and* $(x_0,\cdots,x_n)$.
- *For all* ε *sufficiently small,* $\forall t, k$ *and* $(x_0,\cdots,x_n) \in \Delta_{t+k,k}$, $\Phi'\{(X_0,\cdots,X_{\tau_F}) = (x_0,\cdots,x_n)\} \leq \alpha\delta^t\varepsilon^k$, *where* α *and* δ *are strictly positive constants independent of* ε *and* $(x_0,\cdots,x_n)$.
- *There exists a constant* $l_{r,k} \geq r$, *independent of* ε *and* $(x_0,\cdots,x_n) \in \Delta_{r+k,k}$, *such that* $\sum_{(x_0,\cdots,x_n)\in\Delta_{r+k,k}} |L(x_0,\cdots,x_n) - \gamma| = \Theta(\varepsilon^{l_{r,k}})$ *and, for all* ε *sufficiently small, a strictly positive constant* ν_r^* *independent of* k, ε *and* $(x_0,\cdots,x_n) \in \Delta'_r$ *such that,* $\forall k \geq 0$, $\forall (x_0,\cdots,x_n) \in \Delta_{r+k,k}$, $|L(x_0,\cdots,x_n) - \gamma| \leq \nu_r^*\varepsilon^{l_{r,k}}$.
- *For* ε *sufficiently small,* $\forall t > r$ *and* $(x_0,\cdots,x_n) \in \Delta'_t$, $\gamma \geq |L(x_0,\cdots,x_n) - \gamma| \geq \gamma/2$.

Proof: On Δ'_t, since $\Phi' \in \mathcal{I}$, we can not have more than t failures from a state different than $\mathbf{1}$, then not more than $t+1$ failures on the whole path. After each failure, we can not have more than $N-1$ repairs on $[\tau_F < \tau_{\mathbf{1}}]$. We can not then have more than $t(N-1)$ repairs on the whole path, so, the total number of transitions can not be greater than $tN+1$. Thus

$$|\Delta'_t| \leq |S|^{tN+1}.$$

For the second part, by a similar argument to the one in [11, p. 547], as Γ is finite, for all ε sufficiently small (i.e. $\exists \Upsilon > 0$ such that $\forall \varepsilon < \Upsilon$), for all $(x,y) \in \Gamma$,

$\frac{\boldsymbol{P}(x,y)}{\boldsymbol{P}'(x,y)} = \nu(x,y)\varepsilon^{d(x,y)} + o(\varepsilon^{d(x,y)}) \leq 2\nu(x,y)\varepsilon^{d(x,y)}$, where $d(x,y) \geq 0$ if $x = \mathbf{1}$ or $y \prec x$, $d(x,y) \geq 1$ if $x \neq \mathbf{1}$ and $y \succ x$ (by definition of class $\mathcal{I}$) and $\nu(x,y) > 0$. Then

$$\begin{aligned} L(x_0, \cdots, x_n) &= \prod_{k=0}^{n-1} \frac{\boldsymbol{P}(x_k, x_{k+1})}{\boldsymbol{P}'(x_k, x_{k+1})} \\ &\leq \prod_{k=0}^{n-1} 2\nu(x_k, x_{k+1})\varepsilon^{d(x_k,x_{k+1})}. \end{aligned}$$

Let $\nu' = \max\{2\nu(x,y) : (x,y) \in \Gamma\}$ and $\nu_* = \max(1, \nu')$. We can remark that $\nu_* < \infty$ since $|S| < \infty$. On Δ'_t, as $\sum_{k=0}^{n-1} d(x_k, x_{k+1}) = t$ and $n \leq tN + 1$ as seen in the proof of the first part of this Lemma, then

$$L(x_0, \cdots, x_n) \leq \varepsilon^t \prod_{k=0}^{n-1} 2\nu(x_k, x_{k+1}) \leq \varepsilon^t \nu_*^{tN+1}.$$

If we take $\eta = \nu_*^N$, and $\kappa = \nu_*$ we obtain the desired inequality.

Using the same arguments as in Point 2, there exist functions e and f such that, on $\Delta_{t+k,k} (\subseteq \Delta'_t)$,

$$\Phi'\{(X_0, \cdots, X_{\tau_F}) = (x_0, \cdots, x_n)\} = \prod_{k=0}^{n-1} \boldsymbol{P}'(x_k, x_{k+1}) \leq \prod_{k=0}^{n-1} f(x_k, x_{k+1})\varepsilon^{e(x_k,x_{k+1})} \leq \alpha\delta^t\varepsilon^k,$$

where α and δ are strictly positive and independent of $(x_0, \cdots, x_n)$ and ε.

In an analogous way, for the proof of the fourth part, for all ε sufficiently small, for all $(x_0, \cdots, x_n) \in \Delta_{r+k,k}$, there exist $l(x_0, \cdots, x_n) \geq r$ and $\nu_r(x_0, \cdots, x_n) > 0$ such that $|L(x_0, \cdots, x_n) - \gamma| = \Theta(\varepsilon^{l(x_0,\cdots,x_n)})$ and $|L(x_0, \cdots, x_n) - \gamma| \leq \nu_r(x_0, \cdots, x_n)\varepsilon^{l(x_0,\cdots,x_n)}$. Let $l_{r,k} = \min_{(x_0,\cdots,x_n)\in\Delta_{r+k,k}} l(x_0, \cdots, x_n)$ and $\nu_r^* = \max_{(x_0,\cdots,x_n)\in\Delta'_r} \nu_r^*(x_0, \cdots, x_n)$ ($\nu_r^* < \infty$ because $\Delta'_r < \infty$). Then we obtain the fourth part of the Lemma.

For all ε sufficiently small, for all $t > r$ and $(x_0, \cdots, x_n) \in \Delta'_t$, from point 2 of this Lemma, $\gamma \geq \gamma - L(x_0, \cdots, x_n) \geq \gamma - \kappa\eta^t\varepsilon^t$. Since γ can be written as $\gamma = \sum_{k=0}^{\infty} a_k\varepsilon^{r+k}$ with $a_0 > 0$, we have also $\gamma - \kappa\eta^t\varepsilon^t \geq \frac{\gamma}{2}$ $\forall t > r$ if, for all ε sufficiently small, $\kappa\eta^t\varepsilon^t < \frac{a_0}{3}\varepsilon^r$ $\forall t > r$, which is equivalent to $\varepsilon < \frac{1}{\eta^{t/(t-r)}}\left(\frac{a_0}{3\kappa}\right)^{1/(t-r)}$ $\forall t > r$. But $\inf_{t>r} \frac{1}{\eta^{t/(t-r)}}\left(\frac{a_0}{3\kappa}\right)^{1/(t-r)} > 0$, so we obtain the fifth part of the Lemma. ■

Proof of the Theorem:

Necessary condition: Suppose that there exist $k, m \in \mathbb{N}$ and $(x_0, \cdots, x_n) \in \Delta_{m,k}$ such that $k = 3m/2 - 3s/4 + k'$ with k' strictly positive and $m - k < r$. This means that $\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}=\Theta(\varepsilon^{3m/2-3s/4+k'})$. We have then

$$\begin{aligned}\rho_{\Phi'} &= \sum_{(y_0,\cdots,y_n)\in\Delta} |L(y_0,\cdots,y_n)-\gamma|^3\,\Phi'\{(X_0,\cdots,X_{\tau_F})=(y_0,\cdots,y_n)\}\\ &\geq |L(x_0,\cdots,x_n)-\gamma|^3\,\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}\\ &= \Theta(\varepsilon^{3(m-k)+k})=\Theta(\varepsilon^{3s/2-2k'}).\end{aligned}$$

Thus $\rho_{\Phi'}/\sigma^3_{\Phi'}=\underline{O}(\varepsilon^{-2k'})\to+\infty$.

In the cancellation case, suppose that there exists $k'>0$ such that

$$\sum_{t=r+1}^{\infty}\sum_{(x_0,\cdots,x_n)\in\Delta'_t}\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}=\Theta(\varepsilon^{3s/2-3r-k'}).$$

Then, from Lemma 1, for ε sufficiently small,

$$\begin{aligned}\rho_{\Phi'} &\geq \sum_{t=r+1}^{\infty}\sum_{(x_0,\cdots,x_n)\in\Delta'_t}|L((x_0,\cdots,x_n)-\gamma|^3\,\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}\\ &\geq \left(\frac{\gamma}{2}\right)^3\sum_{t=r+1}^{\infty}\sum_{(x_0,\cdots,x_n)\in\Delta'_t}\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}\\ &= \Theta(\varepsilon^{3s/2-k'}).\end{aligned}$$

In the same way, let us suppose that there exist $k\geq 0$ and $(x_0,\cdots,x_n)\in\Delta_{r+k,k}$ such that $|L(x_0,\cdots,x_n)-\gamma|=\Theta(\varepsilon^l)$ with $l<s/2-k/3$. Let $l'=s/2-k/3-l>0$. Then

$$\begin{aligned}\rho_{\Phi'} &\geq |L(x_0,\cdots,x_n)-\gamma|^3\,\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}\\ &= \Theta(\varepsilon^{3l+k})=\Theta(\varepsilon^{3s/2-3l'}).\end{aligned}$$

In both cases $\rho_{\Phi'}/\sigma^3_{\Phi'}$ is unbounded as $\varepsilon\to 0$.

Sufficient condition: We have to show that

$$\rho_{\Phi'}=\sum_t\sum_{(x_0,\cdots,x_n)\in\Delta'_t}|L(x_0,\cdots,x_n)-\gamma|^3\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}=O(\varepsilon^{3s/2}).$$

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

$$\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\} = \underline{O}(\varepsilon^{3m/2-3s/4})$$

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

$$\begin{aligned}|L(x_0,\cdots,x_n)-\gamma|^3\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\} &= \frac{\Theta(\varepsilon^{3m})}{\Theta(\varepsilon^{3k})}\Theta(\varepsilon^k)\\ &= \frac{\Theta(\varepsilon^{3m})}{\underline{O}(\varepsilon^{2(3m/2-3s/4)})}\\ &= O(\varepsilon^{3s/2}).\end{aligned}$$

As $\sum_{t<r}|\Delta'_t| < +\infty$ and by the first part of Lemma 1,

$$\sum_{t<r}\sum_{(x_0,\cdots,x_n)\in\Delta'_t}|L(x_0,\cdots,x_n)-\gamma|^3\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}=O(\varepsilon^{3s/2}).$$

Consider the case where $s\le 2r$. Since

$$\sum_{t=r}^{\infty}\sum_{(x_0,\cdots,x_n)\in\Delta'_t}\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}\le 1$$

and $\forall t$, $\sum_{(x_0,\cdots,x_n)\in\Delta'_t}\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}\le 1$, Lemma 1 implies that

$$\sum_{t=r}^{\infty}\sum_{(x_0,\cdots,x_n)\in\Delta'_t}|L(x_0,\cdots,x_n)-\gamma|^3\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}$$

$$\begin{aligned}
&\le \sum_{t=r}^{\infty}\sum_{(x_0,\cdots,x_n)\in\Delta'_t}(\gamma^3+3\gamma^2\kappa\eta^t\varepsilon^t+3\gamma\kappa^2\eta^{2t}\varepsilon^{2t}+\kappa^3\eta^{3t}\varepsilon^{3t})\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}\\
&= \gamma^3\sum_{t=r}^{\infty}\sum_{(x_0,\cdots,x_n)\in\Delta'_t}\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}\\
&\quad+\sum_{t=r}^{\infty}(3\gamma^2\kappa\eta^t\varepsilon^t+3\gamma\kappa^2\eta^{2t}\varepsilon^{2t}+\kappa^3\eta^{3t}\varepsilon^{3t})\sum_{(x_0,\cdots,x_n)\in\Delta'_t}\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}\\
&\le \gamma^3+3\gamma^2\kappa\sum_{t=r}^{\infty}(\eta\varepsilon)^t+3\gamma\kappa^2\sum_{t=r}^{\infty}(\eta^2\varepsilon^2)^t+\kappa^3\sum_{t=r}^{\infty}(\eta^3\varepsilon^3)^t\\
&= \Theta(\varepsilon^{3r})+\Theta(\varepsilon^{3r})+\Theta(\varepsilon^{3r})+\Theta(\varepsilon^{3r})=O(\varepsilon^{3s/2})
\end{aligned}$$

because $2r\ge s$.

Consider now the cancellation case. First, by Lemma 1 and the condition on $\Delta_{r+k,k}$,

$$\sum_{(x_0,\cdots,x_n)\in\Delta'_r}|L(x_0,\cdots,x_n)-\gamma|^3\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}$$

$$\begin{aligned}
&= \sum_{k\ge 0}\sum_{(x_0,\cdots,x_n)\in\Delta_{r+k,k}}|L(x_0,\cdots,x_n)-\gamma|^3\Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}\\
&\le \sum_{k\ge 0}\sum_{(x_0,\cdots,x_n)\in\Delta_{r+k,k}}\nu_r^{*3}\varepsilon^{3l_{r,k}}\alpha\delta^r\varepsilon^k\\
&\le \sum_{k\ge 0}\sum_{(x_0,\cdots,x_n)\in\Delta_{r+k,k}}\nu_r^{*3}\alpha\delta^r\varepsilon^{3s/2}\\
&= |\Delta'_r|\nu_r^{*3}\alpha\delta^r\varepsilon^{3s/2}=\Theta(\varepsilon^{3s/2}).
\end{aligned}$$

Second, consider the sum on $\bigcup_{t=r+1}^{\infty} \Delta'_t$. Using point 5 of Lemma 1,

$$\sum_{t=r+1}^{\infty} \sum_{(x_0,\cdots,x_n)\in\Delta'_t} |L(x_0,\cdots,x_n)-\gamma|^3 \Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}$$

$$\leq \gamma^3 \sum_{t=r+1}^{\infty} \sum_{(x_0,\cdots,x_n)\in\Delta'_t} \Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\}$$

for ε sufficiently small. But by assumption

$$\gamma^3 \sum_{t=r+1}^{\infty} \sum_{(x_0,\cdots,x_n)\in\Delta'_t} \Phi'\{(X_0,\cdots,X_{\tau_F})=(x_0,\cdots,x_n)\} = O(\varepsilon^{3s/2}).$$

Thus the normal approximation is bounded. ■

4 Difference between bounded normal approximation and bounded relative error

In the following theorem we prove that the class of measures with bounded normal approximation is included in the class of measures having bounded relative error.

Theorem 5 *Consider a measure $\Phi' \in \mathcal{I}$. If we have bounded normal approximation, we have bounded relative error.*

Proof: In the cancellation case, we have bounded relative error by definition, so the Theorem is true. Then we only have to prove the Theorem in the case without cancellation, i.e. $s \leq 2r$. Suppose that we have a bounded normal approximation. By definition of s there exists at least one $m \geq r$ and one path $(x_0,\cdots,x_n) \in \Delta_m$ such that $\frac{\Phi^2}{\Phi'}\{(X_0,\cdots,X_{\tau_F}) = (x_0,\cdots,x_n)\} = \Theta(\varepsilon^s)$, thus $\Phi'\{(X_0,\cdots,X_{\tau_F}) = (x_0,\cdots,x_n)\} = \Theta(\varepsilon^{2m-s})$. As we have bounded normal approximation, $2m-s \leq 3m/2-3s/4 = 3/4(2m-s)$, then $2m-s \leq 0$. Then $\Phi'\{(X_0,\cdots,X_{\tau_F}) = (x_0,\cdots,x_n)\} = \Theta(1)$ because we have always $2m \geq s$.

Let $(y_0,\cdots,y_n) \in \Delta_r$ and $r' \geq 0$ be such that $\Phi'\{(X_0,\cdots,X_{\tau_F}) = (y_0,\cdots,y_n)\} = \Theta(\varepsilon^{r'})$. By definition of r,

$$2r \geq 2r - r' \geq s = 2m \geq 2r$$

with the previous m, then

$$s = 2r.$$

Suppose that we do not have bounded relative error. There exists $(z_0, \cdots, z_n) \in \Delta$, with $\Phi\{(X_0, \cdots, X_{\tau_F}) = (z_0, \cdots, z_n)\} = \Theta(\varepsilon^l)$ and $\Phi'\{(X_0, \cdots, X_{\tau_F}) = (z_0, \cdots, z_n)\} = \Theta(\varepsilon^{l'})$, such that $l' > 2l - 2r$, which means that $l' > 2l - s$. But by definition of s, $s \leq 2l - l'$. We have then proved the theorem. ■

We can find a system example with bounded relative error, but without bounded normal approximation. Suppose that $C = 2$, $n_1 = n_2 = 2$ and that the system is operational if at least two components are operational. Let the transitions of the DTMC of this system be represented by the transitions on Figure 1, where the failed states are shaded, and state i, j means that there are i components up of type 1 and j components up of type 2. If we use as importance sampling scheme Bias1 failure biasing, with the new probabilities described in Figure 2, then we have bounded

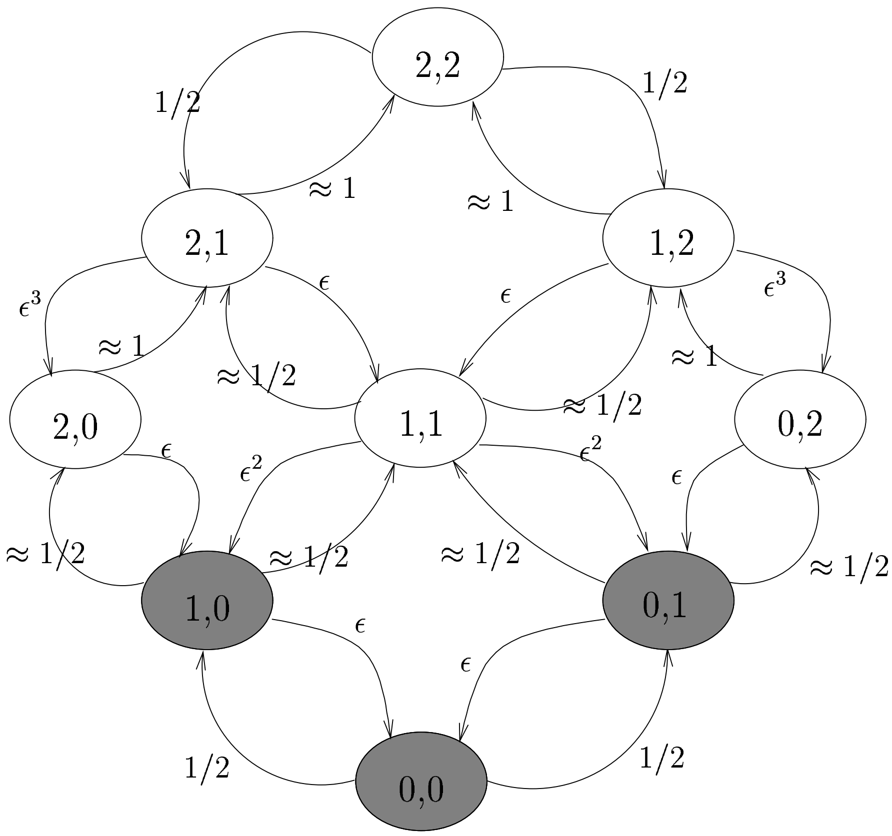

Figure 1: Transitions of system I

relative error, but not bounded normal approximation. This means that, even if we have bounded relative error, the confidence interval will have poor coverage for high reliability values. For this example,

$$E_\Phi(1_{[\tau_F<\tau_{\mathbf{1}}]}) = 2\varepsilon^3 + o(\varepsilon^3),$$

$$E_{\Phi'}(1_{[\tau_F<\tau_{\mathbf{1}}]}L^2) = \frac{5}{\rho_0^2}\varepsilon^6 + o(\varepsilon^6)$$

and

$$\rho_{\Phi'} = \frac{1}{\rho_0^4}\varepsilon^8 + o(\varepsilon^8).$$

For this system, Bias1 failure biasing and failure distance biasing importance sam-

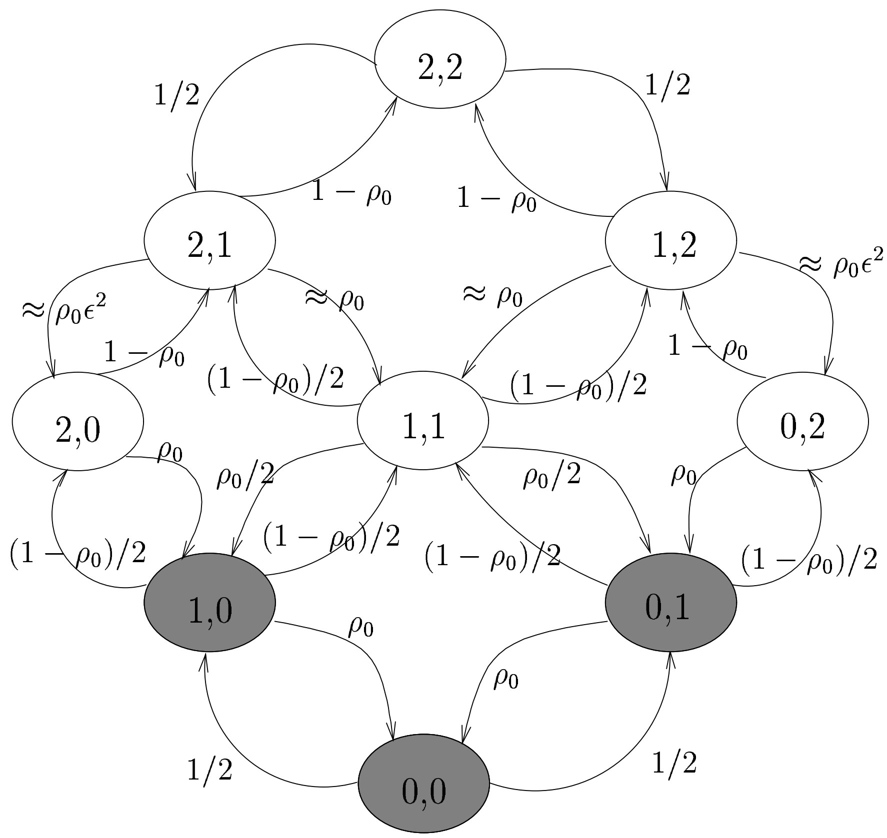

Figure 2: Bias1 failure biasing importance sampling transitions of system I

pling measures give the same values for the moments, so, this is also an example of the failure distance biasing importance sampling technique giving bounded relative error but not bounded normal approximation.

As a numerical illustration, if we make 10^4 different estimations with $\varepsilon = 10^{-10}$, $I = 10^3$ and using the asymptotic expansions of γ and $\sigma^2_{\Phi'}$ to consider the confidence interval at confidence level 0.95, we can see that the true value is contained in the confidence interval 94.6% of the time for balanced failure biasing and 98.9% for Bias1 failure biasing. Thus, with Bias1 failure biasing, the coverage is significantly better than the expected 95%. We can imagine that there may be systems producing inverse effects.

We can also exhibit systems (with at least three types of components) with the same property for the Bias2 failure biasing technique.

Theorem 6 *Balanced failure biasing has the property of bounded normal approximation. Bias1 failure biasing, Bias2 failure biasing and failure distance biasing do not always have the bounded normal approximation property. Nevertheless, for balanced systems (i.e. systems for which failure transitions from state* **1** *have order 1 probabilities whereas failures from other states have probabilities of the same order of* ε*), all the above methods give bounded normal approximation.*

Proof: The proof that balanced failure biasing give bounded normal approximation results directly from the necessary and sufficient condition given in Theorem 4 and from the fact that $\Phi'\{(X_0, \cdots, X_{\tau_F}) = (x_0, \cdots, x_n)\} = \Theta(1)$ (i.e. $k = 0$, where k is the same as in the definition of $\Delta_{m,k}$) for all paths to failure $(x_0, \cdots, x_n) \in \Delta$, because each probability in matrix $\boldsymbol{P'}$ is in $\Theta(1)$ [13]. As a matter of fact, the condition of Theorem 4 on Δ'_t with $t < r$ is verified because $k = 0$. Moreover, in the cancellation case, $\Delta'_t = \Delta_{t,0} = \emptyset$ for $t > r$ and the cancellation on $\Delta'_r = \Delta_{r,0}$ is at least of order $l_{r,0} \geq s/2$ (otherwise $\sigma^2_{\Phi'} = \underline{O}(\varepsilon^{2r})$).The same arguments work in the case of balanced systems. Counter-examples for Bias1 failure biasing and the distance-based technique are given above. A counter-example for Bias2 failure biasing can also be built in a similar manner. ■

5 Conclusion

The objective of this paper is to define the concept of bounded normal approximation and to emphasize its importance in the context of the evaluation of dependability measures using Markov models. Then we give necessary and sufficient conditions to obtain bounded normal approximation in simulations of highly reliable Markovian systems. Up to now, literature has focused on bounded relative error. A good importance sampling measure should verify both properties. In this context, we show that bounded normal approximation always implies bounded relative error. We then give examples to show that the reverse implication is not true.

Balanced failure biasing, that is known to give bounded relative error for the large class of systems described, always gives bounded normal approximation. The other three commonly used biasing schemes - Bias1, Bias2 and failure distance biasing - are known to give bounded relative error when the system is balanced; these schemes also give bounded normal approximation when the system is balanced. In short, all the scheme-system combinations that are likely to occur in practice:

- balanced failure biasing for balanced or unbalanced systems,
- the other three schemes for balanced systems,

have the bounded normal approximation property.

Acknowledgements

I would like to thank the referee for his numerous constructive comments and editorial corrections. I also wish to thank a previous referee who drew my attention on the recent version of the Berry-Esseen Theorem which uses the estimated standard deviation, and Paul Glasserman and Gerardo Rubino for their remarks.

References

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

[2] J. A. Carrasco. Failure distance based on simulation of repairable fault tolerant systems. *Proceedings of the 5th International Conference on Modeling Techniques and Tools for Computer Performance Evaluation*, pages 351–365, 1992.

[3] W. Feller. *An Introduction to Probability Theory and its Applications*, volume II. John Wiley and Sons, 1966. second edition.

[4] A. Goyal, L. Lavenberg, and K. Trivedi. Probabilistic Modeling of Computer System Availability. *Annals of Operations Research*, 8:285–306, 1987.

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

[6] P. Hall. Edgeworth expansion for Student's t statistic under minimal moment conditions. *The Annals of Probability*, 15(3):920–931, 1987.

[7] P. Hall. *The Bootstrap and Edgeworth Expansion.* Springer-Verlag, New York, 1992.

[8] J. M. Hammersley and D. C. Handscomb. *Monte Carlo Methods.* Methuen, London, 1964.

[9] E. E. Lewis and F. Böhm. Monte Carlo Simulation of Markov Unreliability Models. *Nuclear Engineering and Design*, 77:49–62, 1984.

[10] R.R. Muntz, E. de Souza e Silva, and A. Goyal. Bounding Availability of Repairable Computer Systems. *IEEE Transactions on Computers*, 38(12):1714–1723, 1989.

[11] M. K. Nakayama. Asymptotics of Likelihood Ratio Derivatives Estimators in Simulations of Highly Reliable Markovian Systems. *Management Science*, 41(3):524–554, March 1995.

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

[13] P. Shahabuddin. Importance Sampling for the Simulation of Highly Reliable Markovian Systems. *Management Science*, 40(3):333–352, March 1994.