Proceedings of the 2005 Winter Simulation Conference
M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, eds.

NEW MEASURES OF ROBUSTNESS IN RARE EVENT SIMULATION

H. Cancela

Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay

G. Rubino and B. Tuffin

IRISA-INRIA
Campus Universitaire de Beaulieu
35042 Rennes Cedex, France

ABSTRACT

Rare event simulation requires acceleration techniques in order to i) observe the rare event and ii) obtain a valid and small confidence interval for the expected value. A "good" estimator has to be robust when rarity increases. This paper aims at studying robustness measures, the standard ones in the literature being Bounded Relative Error and Bounded Normal Approximation. By considering the problem of estimating the reliability of a static model for which simulation time per run is the critical issue, we show that actually those measures do not validate the satisfying behavior of some techniques. We thus define Bounded Relative Efficiency and generalized bounded normal approximation properties of the two previous measures in order to encompass the simulation time. We also illustrate how a user can have a look at the coverage of the resulting confidence interval by using the so-called coverage function.

1 INTRODUCTION

In order to understand the behavior and dynamics of systems, stochastic modeling has been seen as a powerful tool, with applications in various fields such as biology, medicine, computer science, telecommunications... The solution methods available to analyze those models depend on their degree of abstraction: analytical techniques are usually only applicable when quite stringent assumptions are verified such as Markov property for instance. If no necessary assumption is imposed, the users have to turn to numerical solutions. Another requirement for both analytical and numerical techniques is that the model is of small to moderate size (or encompasses symmetries) so that the time to get the results remains feasible. If those two requirements are not met, Monte Carlo simulation becomes the method of choice.

Another issue is that, in many cases, the properties of interest depend critically on the occurrence of a rare event. This happens for instance in communication networks, where a traditional measure of performance is the packet loss probability having an incidence on quality of service. Losses are generally due to buffer overfilling and may be computed from queuing theory tools. Another field of interest is dependability analysis. For instance, we may be interested in computing the probability of not being able to connect two terminal sites of a network, due to failures in the links. Usually, these events have very small probabilities, smaller than 10^{-9} in actual practice. Nuclear plants reliability analysis is a similar critical application.

When we find a rare event situation, standard simulation techniques meet important difficulties, as the low probability of the interesting states makes it very improbable to observe them in a random sample of the evolution of the system. This leads to very poor precision in the estimation of the target measures, and increases the probability of non meaningful experiments (such as never observing the event of interest). There has been much research in alternative techniques, which can improve the precision of the estimation. Most of these methods are usually classified within the class of variance reduction techniques, as they strive to give estimators for the target measure having the same mean value but smaller variance than the standard Monte Carlo estimator (Fishman 1997). This improved precision is in general attained at the cost of employing a more complex algorithm, which leads in many cases (but not necessarily) to increasing the computational time. Other methods have the same precision per replication as standard Monte Carlo, but with lower computational costs.

As there are many techniques offering different tradeoffs, two immediate questions are i) is the method robust as rarity increases? ii) what is the most appropriate one? To answer those questions, a purely empirical approach, used in some of the first papers in the area, consists of observing the efficiency of each

method (defined as the product of the variance by the computational time) in comparison with the standard Monte Carlo technique, which serves as a base point, over a test set. This idea has many problems, as it is difficult to extrapolate the results for other systems not included in the test set, and to obtain useful insights to design alternative methods. A better possibility is the analytical study of the simulation methods. In particular, there has been a line of research of the asymptotic behavior of rare event simulation estimators when the rarity of the events goes to 0, which has led to define new concepts such as bounded relative error, asymptotic optimality (both concerning the size of the confidence interval relatively to the considered rare event), and more recently bounded normal approximation (looking at the coverage of the confidence interval). These concepts focus on the precision attained and the robustness of the simulation estimators (both are important features, see for example the discussion in (Heegaard 1998)), but do not take into account the computational times associated with them. We show in this paper that those properties are not defined generally enough to encompass some classes of estimators actually exhibiting a very nice behavior. This is illustrated on the problem of estimating the source-terminal reliability of a network. We therefore generalize the above mentioned properties by incorporating the average simulation time to get a single measure. This leads to the definitions of bounded relative efficiency and generalized bounded normal approximation. Another concept that we introduce in this paper is the use of the coverage function (Schruben 1980) applied to rare event estimators to exactly see in practice if the coverage of the estimator is robust with respect to the rarity of the event.

The paper is organized as follows. In Section 2, we recall the properties from the existing literature: bounded relative error, asymptotic optimality, and bounded normal approximation. In Section 3 we first present a simulation method for estimation the static source-terminal reliability in a network that will be used as the illustration throughout the paper. We show that, despite being efficient, this method does not verify the above traditional properties. We thus propose a desirable property for a simulation method, called Bounded Relative Efficiency, which corresponds to the situation where a given relative error can be obtained with constant computational effort even when the probability of the event of interest goes to 0. We also provide for our illustrative simulation method a sufficient condition for holding Bounded Relative Efficiency, implying its robust behavior. Section 4 then similarly generalizes bounded normal approximation property, providing a bound of error coverage in average for a given simulation time. Again, a condition is provided for the static reliability estimator. This last property giving only a *sufficient* condition for ensuring the coverage of the confidence interval, we look in Section 5 at the effective coverage by using the so-called coverage function. Section 6 presents numerical examples illustrating our results and Section 7 concludes the paper and gives directions for future research.

2 PROPERTIES OF RARE EVENT ESTIMATORS

Throughout the rest of the paper, we aim at estimating the probability γ of a rare event. Since we wish to study the robustness of estimators as rarity increases, we introduce a rarity parameter ε characterizing the rare event and such that as $\varepsilon \to 0$, $\gamma \to 0$. This parameter may have different interpretations depending on the context of use. In reliability models for instance, ε may represent a maximum failure rate of a component in the case of dynamic models (Shahabuddin 1994), or the reliability of a component in the case of static models. In queuing performance evaluation models, ε may be chosen as $1/B$ where B is the buffer size, so that the buffer overflow probability $\gamma \to 0$ as $\varepsilon \to 0$ (Juneja 1993, Heegaard 1998).

2.1 Bounded Relative Error

Let us consider an unbiased estimator $\hat{\gamma}_n$ of γ built from a sample having size n. Bounded Relative Error (BRErr) has been defined in (Shahabuddin 1994) in order to state if the half-width confidence interval divided by γ is bounded as ε tends to 0 (for a fixed sample size n). This asserts the robustness of the estimation, meaning that the relative error is not sensitive to the rarity of the event. Formally,

Definition 1 *Let σ_n^2 denote the variance of $\hat{\gamma}_n$, $\sigma_n = \sqrt{\sigma_n^2}$ and let z_δ denote the $1 - \delta/2$ quantile of the standard normal distribution ($z_\delta = \mathcal{N}^{-1}(1 - \delta/2)$ where $\mathcal{N}$ is the standard normal distribution). Then the relative error RErr is defined by*

$$RErr = z_\delta \frac{\sigma_n}{\gamma}. \tag{1}$$

We say that we have a bounded relative error (BRErr) if RErr remains bounded as $\varepsilon \to 0$.

This property has been further studied in detail in several articles (Heidelberger 1995, Heidelberger, Shahabuddin, and Nicola 1994, Nakayama 1996), mainly in the context of highly reliable Markovian systems. Necessary and sufficient conditions for verifying the property are provided in those papers.

2.2 Asymptotic optimality

Asymptotic optimality has been widely used in queuing applications, for a special class of simulation methods called importance sampling. Importance sampling consists of modifying the probability measure of the system under study before simulating it: if $\gamma = E_f[g(X)]$ is the expectation of random variable $g(X)$ under probability measure f, then $\gamma = \int g(x)f(x)dx = \int g(x)L(x)f^*(x)dx = E_{f^*}[g(X)L(X)]$ where $L(x) = f(x)/f^*(x)$ is called the likelihood ratio (assuming $f^* > 0$ if $fg > 0$). In other words, γ is also the expectation of $g(X)L(X)$ under probability measure f^*. Denoting $\hat{\gamma}_n = n^{-1}\sum_{i=1}^n g(X_i)$ the standard estimator of γ where the sample is $X_1, \cdots, X_n$, the importance sampling estimator of γ corresponding to the new measure f^* is $\hat{\gamma}_n^{IS} = n^{-1}\sum_{i=1}^n g(X_i)L(X_i)$. Then, if $L \ll 1$ the variance of $\hat{\gamma}_n^{IS}$ verifies $\mathrm{Var}(\hat{\gamma}_n^{IS}) \ll \mathrm{Var}(\hat{\gamma}_n)$.

Definition 2 *An importance sampling estimator $\hat{\gamma}_n^{IS}$ is called asymptotically optimal, if*

$$\lim_{\varepsilon \to 0} \frac{\ln E_{f^*}[g(X)^2 L(X)^2]}{\ln \gamma} = 2.$$

Note that the quantity under limit is always positive and less than or equal to 2.

Basically, this property means that when $\varepsilon \to 0$ the variance of $\hat{\gamma}_n^{IS}$ goes to zero as well. However, what interests us here is the relative error. In (Sandmann 2004), it is proved that asymptotic optimality is a necessary but not sufficient condition to BRErr.

2.3 Bounded Normal Approximation

Whereas the two previous properties deal with the variance of the estimator to maintain as small as possible the relative size of the confidence interval, an important remaining question is whether or nor the coverage of this confidence interval remains bounded as $\varepsilon \to 0$.

Bounded Normal Approximation (BNA) (Tuffin 1999) ensures that the Gaussian approximation, and thus the confidence interval coverage, remains valid as ε tends to 0. It is based on the Berry-Esseen Theorem which states that if ϱ is the third absolute moment of each of the n i.i.d. random variables X_i (and σ^2 the variance), $\mathcal{N}$ the standard normal distribution, $\hat{\gamma}_n = n^{-1}\sum_{i=1}^n X_i$, $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^I (X_i - \hat{\gamma}_n)^2$ and F_n the distribution of the centered and normalized sum $(\hat{\gamma}_n - \gamma)/\hat{\sigma}_n$, then there exists an absolute constant $a > 0$ such that, for each x and I

$$|F_n(x) - \mathcal{N}(x)| \leq \frac{a\varrho}{\sigma^3\sqrt{n}}. \qquad (2)$$

Definition 3 *We say that $\hat{\gamma}_n$ verifies Bounded Normal Approximation if ϱ/σ^3 remains bounded as $\varepsilon \to 0$.*

If the estimator enjoys this property, only a fixed number of iterations is required to obtain a confidence interval having a fixed error no matter how rarely failures occur.

In (Tuffin 2004), it is shown that, for Markovian reliability models, BNA implies that the estimation of the variance is asymptotically correct, implying in turn BRErr, which implies the γ is well-estimated, but that none of the converse implications is verified in full generality. A refinement of the necessary and sufficient condition for BNA is also provided in (Tuffin 2004).

3 DEFINITION OF BOUNDED RELATIVE EFFICIENCY

3.1 Need for extending the current properties

Consider the problem of evaluating the reliability of a "static" (time is not an explicit variable) stochastic model of a complex system by Monte Carlo. To be specific, consider a standard network reliability problem: we are given an undirected graph G representing a communication network where nodes are perfect but links (edges) can fail (they can be either operational or completely down), two fixed nodes, s and t, and we want to quantify the capacity of the network to support the communications between these two selected nodes. Edges are supposed to fail independently, and we know the (elementary) reliability r_i of each edge i (r_i is the probability that edge i is working). The random set of operational edges defines a random subgraph G' of G. The target is the network reliability R, the probability that nodes s and t belong to the same connected component of G'.

The computation of R is NP-hard, and it is out of scope today for even moderate graph sizes (having, say, several dozens of nodes and links (Rubino 1998)). Estimating R using a standard Monte Carlo method consists of building n copies of G', simply counting in how many of those the selected nodes can communicate and dividing this number by n. This ratio is an unbiased estimator of R; its variance is $R(1-R)/n$. The cost of building a copy of G' following the standard approach is $\Theta(M)$ if M is the number of links in G, and the average cost in time of checking if s and t are connected in a subgraph of G (that is, the cost of running a Depth First Search procedure) is also $\Theta(M)$. The usual situation is that the reliabilities of the lines are high, leading to a rare event situation ($1 - R \approx 0$).

In (Khadiri and Rubino 1996), a different estimator of R is proposed, having some interesting properties.

To use it, we need to build a set of elementary paths connecting nodes s and t, such that any pair of paths share only nodes s and t. Let this set be $\mathcal{P} = \{P_1, P_2, \cdots, P_H\}$, and call π_h the event "all links of path P_h work". Denote by p_h the probability of π_h, that is,

$$p_h = \Pr[\pi_h] = \prod_{i \in P_h} r_i.$$

Consider an infinite sequence of independent copies of G' and let F be the random variable "first element in the sequence where every path in $\mathcal{P}$ has at least one link that does not work". See that

$$\Pr[F = 1] = q = \prod_{h=1}^{H} (1 - p_h),$$

and in general, for any $n \geq 1$,

$$\Pr[F = n] = (1 - q)^{n-1} q.$$

Then, on the average, we have to wait for $\mathrm{E}(F) = 1/q$ samples of G' to find one such that no path in $\mathcal{P}$ connects s and t. If the links are highly reliable, then q will be small and $\mathrm{E}(F)$ large. It was then proven in (Khadiri and Rubino 1996) that the following intuitive idea works. We first sample from the geometric distribution of F. Call f the obtained value. The estimator of R is then built assuming that in the first $f-1$ copies of G' nodes s and t are connected (saving a significant amount of computations as the reliability gets close to one). It remains to know if they are connected in the fth copy or not. To build that fth copy, we must sample the states of the lines in the network, *conditioned* to the fact that each of the paths on $\mathcal{P}$ contains *at least* a failed component. The problem reduces to sampling the states of the edges in a path, knowing that at least one of them is down; once this is done for the H paths, the rest of the links in the network are sampled independently, using their original reliabilities.

Let path P_h be $P_h = (i_{h,1}, i_{h,2}, \cdots, i_{h,K_h})$ and let $c_{h,k} = r_{i_{h,1}} r_{i_{h,2}} \cdots r_{i_{h,k}}$ be the probability that the first k edges in P_h (in some arbitrary and fixed order) are all up, for $1 \leq k \leq K_h$. Then, define a random variable W_h on the set of integers $\{1, 2, \cdots, K_h\}$ with distribution

$$\Pr[W_h = k] = \frac{(1 - r_{i_{h,k}}) c_{h,k-1}}{1 - c_{h,K_h}},$$

where $c_{h,0} = 1$. It can be then shown that W_h has the distribution of the index of the random variable "first failed edge of P_h knowing that there is at least one failed edge". To sample the state of links $i_{h,1}, i_{h,2}, \cdots, i_{h,K_h}$ we just sample W_h; if the obtained value is w_h, links $i_{h,1}, i_{h,2}, \cdots, i_{h,w_h-1}$ are set to 'up', link i_{h,w_h} is set to 'down', and the states of the remaining links i in the path (from position w_h+1 to position K_h) are sampled from the original Bernoulli distribution with parameter r_i.

Consider the average cost in building n copies of G' using the previously described approach. We will need, on the average, nq samples from the geometric distribution. For each of these nq cases where we must sample the conditional state of the links in the network, we need to sample from $W_1, \cdots, W_H$, then to sample the states of a subset of the whole graph, which has average cost in $O(M)$. This leads to an average global cost in time of the form $O(nq(M + K))$, where $K = K_1 + \cdots + K_H$. Observe that the variance of the estimator is $R(1-R)/n$, because as stated in (Khadiri and Rubino 1996), we are in fact building the standard estimator in a more efficient way.

Introduce the rarity parameter ε by assuming that, $\forall i$, there exist two reals $a_i, b_i > 0$ such that $r_i = 1 - a_i \varepsilon^{b_i}$. It is straightforward to verify that the unreliability $\gamma = 1 - R \to 0$ as $\varepsilon \to 0$. Let $\hat{\gamma}$ be the above estimator of the unreliability.

The Relative Error of this method is

$$\Theta(\sqrt{\gamma(1-\gamma)}/\gamma) = \Theta(1/\sqrt{\gamma}) \to \infty$$

as $\varepsilon \to 0$. Nevertheless, as the per-replication computational time decreases with ε, this should be also considered in the asymptotic efficiency of the estimator.

3.2 Definition

For a fixed sample size, we thus define the Bounded (Relative) Efficiency. It basically gives the (relative) variance of an estimator obtained during a given simulation time. Indeed, an estimator A yielding a smaller variance than an estimator B for the same number n of replications may require a larger computational time in order to obtain one replication. The efficiency looks at the variance obtained for a given simulation time since a quicker estimator will run more replications.

Definition 4 *Let $\hat{\gamma}_n$ be an estimator of γ built using n replications and σ_n^2 its variance (possibly dependent). Let t_n be the average simulation time to get those n replications. The relative efficiency of $\hat{\gamma}_n$ is given by*

$$REff = \frac{\gamma^2}{\sigma_n^2 t_n}.$$

We will say that $\hat{\gamma}_n$ has bounded relative efficiency (BREff) if there exists a constant $d > 0$ such that REff is minored by d for all ε.

In the case of independent replications, $t_n = nt$ and $\sigma_n^2 = \sigma^2/n$ with t and σ^2 respectively the average time and variance for a single replication. Therefore the efficiency is REff $= \gamma^2/(\sigma^2 t)$, independent of the sample size n.

Note again that the average per-replication simulation time may vary with ε (as well as σ^2 and γ).

3.3 Sufficient condition for BREff on our static reliability estimator

Returning to our unreliability estimation problem, using n *virtual* replications, $\sigma_n^2 = \gamma(1-\gamma)/n$ and $t_n = O(nq(M+K)) = O(nqM)$, since $K \leq M$. We can thus write $t_n = O(nq)$. The efficiency of this approach is then $\underline{O}(\gamma^2 n/(\gamma(1-\gamma)nq)) = \underline{O}(\gamma/q)$ where a function $f(\varepsilon) = \underline{O}(g(\varepsilon))$ if there exist $d_1 > 0$ such that $f(\varepsilon)/g(\varepsilon) \geq d_1$ for all ε sufficiently small.

Using standard representations of R in terms of minpaths or mincuts (see (Rubino 1998)) we know that $\gamma = \Theta(\varepsilon^r)$ for some real $r > 0$ (if every b_i is an integer, then r is an integer as well). Let $\mathcal{C}$ denote the set of mincuts with probability $\Theta(\varepsilon^r)$ (the probability of a mincut is the probability that all its components are down). We also say that mincuts in $\mathcal{C}$ have "order" r. Every other mincut not in $\mathcal{C}$ has probability $\Theta(\varepsilon^{r'})$ (or order r').

A sufficient condition for Bounded Relative Efficiency is then the following.

Theorem 1 *$\forall P_h \in \mathcal{P}$, let $P_h = (i_{h,1}, \cdots, i_{h,K_h})$ and $b_h = \min_{1\leq k\leq K_h} b_{i_{h,k}}$ the order of the most reliable edge of P_h. The estimator $\hat{\gamma}$ of the static unreliability described in previous section verifies Bounded Relative Efficiency if $\sum_{h=1}^{H} b_h \geq r$.*

Proof: Let $a_h = \sum_{k:b_{i_{h,k}}=b_h} a_{i_{h,k}}$. We have

$$\begin{aligned} q &= \prod_{h=1}^{H}(1-p_h) = \prod_{h=1}^{H}\left(1 - \prod_{k=1}^{K_h}\left(1 - a_{i_{h,k}}\varepsilon^{b_{i_{h,k}}}\right)\right) \\ &= \prod_{h=1}^{H}\Theta(a_h\varepsilon^{b_h}) = (\prod_{h=1}^{H} a_h)\Theta(\varepsilon^{\sum_{h=1}^{H} b_h}). \end{aligned}$$

Then REff $= \underline{O}(\gamma/q) = \underline{O}(\varepsilon^{r-\sum_{h=1}^{H} b_h}) = \underline{O}(1)$ (meaning that BREff is verified) if $\sum_{h=1}^{H} b_h \geq r$. □

The method is thus robust as $\varepsilon \to 0$ whereas BRErr is never satisfied.

4 GENERALIZED BOUNDED NORMAL APPROXIMATION

Similarly to the BRErr property, BNA does not deal with the computational time per run. A natural question is then to see whether BNA could also be generalized. Recall that Equation (2) bounds from above the distance between the (normalized and centered) empirical distribution and the Gaussian law by $a\varrho/(\sigma^3\sqrt{n})$ when using n independent runs. The generalization of BNA rather considers the (average) distance between the two distributions for a given simulation time instead of a number of runs (by using the average computational time per run). This leads to the following definition:

Definition 5 *Let $n(T)$ be the average number of runs for a given simulation time T. We say that the estimator $\hat{\gamma}$ verifies Generalized Bounded Normal Approximation (GBNA) if $\varrho/(\sigma^3\sqrt{n(T)})$ remains bounded as $\varepsilon \to 0$, or equivalently if $\varrho\sqrt{t_1}/\sigma^3$ remains bounded as $\varepsilon \to 0$ since $T = n(T)t_1$ (with t_1 the average simulation time for a single run).*

This definition says that if GBNA is verified, the coverage of the confidence interval is robust as the rarity increases, *when considering a fixed simulation time.*

We then have the following theorem saying that, at least for the static reliability analysis problem, GBNA implies $BREff$, and what is more, that actually both properties are equivalent.

Theorem 2 *For our static reliability estimator, GBNA is verified if and only if BREff is verified.*

Proof: Since $\sigma^2 \approx \gamma$, BREff is verified if and only if $\sigma^2 t_1/\gamma^2 \approx t_1/\gamma$ is bounded as $\varepsilon \to 0$.

But $\varrho \approx \gamma$ also, thus GBNA is verified if and only if $\sqrt{t_1}/\sigma^3$ is bounded as $\varepsilon \to 0$, which is equivalent to BREff. □

5 COVERAGE ERROR

This section deals with a way to (exactly) look at the coverage of an estimator. Whereas GBNA bounds the distance between the empiric and Gaussian distributions (and therefore bounds from above the coverage error of the confidence interval), there indeed exists a way to directly look at the coverage error of the confidence interval (instead of bounding it), based on the seminal paper from L.W. Schruben (Schruben 1980). Assume with full generality that a confidence interval $R(\eta, \mathbb{X})$ is constructed for the estimation of parameter γ, at confidence level η with (random) data $\mathbb{X}$. Given the randomness of the data, if the interval estimation is based on true assumptions we have $\Pr[\gamma \in R(\eta, \mathbb{X})] = \eta$. If we define $\eta^* = \inf\{\eta \in [0,1] : \gamma \in R(\eta, \mathbb{X})\}$, then, η^* should be uniformly distributed:

$$F_{\eta^*}(\eta) = \Pr[\eta^* \leq \eta] = \eta.$$

For each *desired* coverage level η, $F_{\eta^*}(\eta)$ is the *actual* coverage level. If $F_{\eta^*}(\eta) < \eta$, the coverage is overstated and may lead to wrong conclusions, while the case $F_{\eta^*}(\eta) > \eta$ means that the desired coverage could

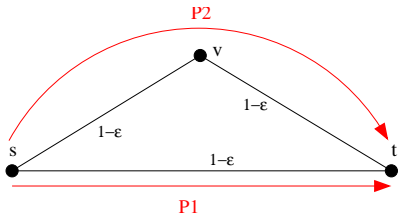

Figure 1: A simple topology.

have been reached at less cost, so that the method is not efficient.

In practice, the distribution of η^* is determined by using I independent data sets $\mathbb{X}_i$ $(1 \leq i \leq I)$ and computing the corresponding values of η_i^*. From these values, the empirical distribution of η^* can be built.

Turning back to rare event estimation, let us now define the Coverage Error.

Definition 6 *Let $n(T)$ be the average number of runs for a given simulation time T. Let $\mathbb{X}_{n(T)}$ be the $n(T)$ data used for an estimation $\hat{\gamma}$ of γ, and $\eta^*_{n(T)}$ be the random variable defined as $\eta^*_{n(T)} = \inf\{\eta \in [0,1] : \gamma \in R(\eta, \mathbb{X}_{n(T)})\}$ (where $R(\cdot,\cdot)$ is usually the centered interval using the empiric standard deviation estimator). The Coverage Error function is defined by $CErr(\eta) = |F_{\eta^*_{n(T)}}(\eta) - \eta|$ which depends on ε.*

Note that the Coverage Error is always bounded since $\sup_{\eta\in[0,1]} |F_{\eta^*_{n(T)}}(\eta) - \eta| \leq 1$ for all $\varepsilon > 0$. Studying this function in terms of ε, as well as Kolmogorov-Smirnov statistic $\sup_{\eta\in[0,1]} |F_{\eta^*_{n(T)}}(\eta) - \eta|$, might nevertheless be of interest.

In the next section, we are going to study the evolution of the coverage function as $\varepsilon \to 0$.

6 EXAMPLES

Consider as an illustration the static reliability estimation problem when using the estimator described in Section 3.1. Consider the topology of Figure 1 when looking at the connectivity between nodes s and t.

6.1 Small illustrative problem

This very simple topology will enable us to derive explicit expressions of the considered metrics and to check if the properties are verified or not. Let $r_i = 1 - \varepsilon$ be the reliability of each link i of the graph, and P_1 and P_2 be the disjoint paths described on Figure 1. It can easily be verified that the unreliability between nodes s and t is $\gamma = 1 - R = \varepsilon^3 + 2\varepsilon^2(1-\varepsilon) \approx 2\varepsilon^2$. Moreover, the variance for a single estimation is given by $\sigma^2 = \gamma(1-\gamma) \approx 2\varepsilon^2$. Therefore, $\sigma/\gamma \approx 1/(\sqrt{2}\varepsilon)$, meaning that we do not have BRErr as $\varepsilon \to 0$, when using the naive implementation of crude Monte Carlo.

On the other hand, the probabilities that all links of P_1 and P_2 work are $p_1 = 1 - \varepsilon$ and $p_2 = (1-\varepsilon)^2$ respectively. Thus the probability that at least one link does not work on each path is $q = (1-p_1)(1-p_2) \approx 2\varepsilon^2$. The simulation time is proportional to the parameter q of the geometric law giving the first time of a failure on the disjoint paths t_n proportional to q, so that REff $= \Theta(\frac{\gamma^2}{\sigma^2 q})$ is bounded. BREff is therefore actually verified.

Also

$$\begin{aligned}\rho &= \mathrm{E}\left(1_{\{s \text{ and } t \text{ not connected}\}} - \gamma\right)\\ &= (1-\gamma)^3(\varepsilon^3 + 2\varepsilon^2(1-\varepsilon))\\ &\quad +\gamma^3((1-\varepsilon)^3 + 3(1-\varepsilon)^2\varepsilon + \varepsilon^2(1-\varepsilon))\\ &\approx 2\varepsilon^2 \approx \gamma.\end{aligned}$$

Since $\sigma^3 \approx 2\sqrt{2}\varepsilon^3$ and $n(T)$ is inversely proportional to $q \approx 2\varepsilon^2$, GBNA is also verified.

Let us now look at the numerical values that can be obtained in practice. The first columns of Table 1 display the estimated value, the confidence interval (at confidence level 95% the Relative Error observed in practice when the number of replications is fixed to $n = 10^4$ and $\varepsilon \to 0$. It can be immediately observed that, for a fixed number of iterations, the quality of the method is getting worse, until the rare event is not observed anymore (therefore exhibiting the same kind of behavior than crude Monte Carlo would). The last column displays the Kolmogorov-Smirnov statistic (the supremum over the coverage error) when using 500 estimations for the empirical coverage function. It becomes equal to 1 as soon as the rare event is not observed anymore. The coverage function for the different values of ε is displayed in Figure 2. The curve for $\varepsilon = 0.9999$ is not displayed since all the mass of the empirical distribution is at value 1. The coverage quality is observed to degrade as $\varepsilon \to 0$.

Table 1: Results on the simple topology, with a number of replications fixed to $n = 10^4$

r_i $\forall i$	Est.	Conf. interval	RErr	KS stat.
0.5	3.779e-01	(3.684e-01,3.874e-01)	2.515e-02	5.671e-02
0.9	1.901e-02	(1.899e-02,1.903e-02)	1.049e-03	4.861e-02
0.95	4.100e-03	(2.848e-03,5.352e-03)	3.055e-01	5.269e-02
0.99	2.000e-04	(-7.717e-05,4.772e-04)	1.386e+00	2.544e-01
0.995	0	(0, 0)	—	1
0.999	0	(0, 0)	—	1
0.9999	0	(0, 0)	—	1

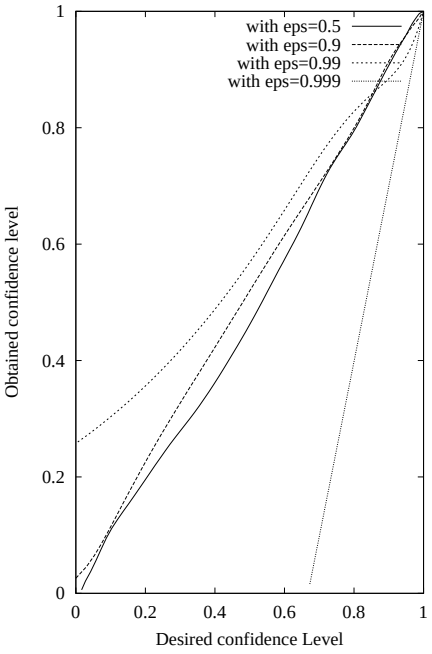

Figure 2: Coverage function for different values of ε and the simple topology. The curve for $\varepsilon = 0.9999$ is not displayed since all the all mass of the empirical distribution is at 1.

In Table 2 the same kind of results are displayed, but using the average number of replications for a simulation time fixed to $T = 10$ seconds on our computer, leading to different values of $n(T)$. It can be observed that the relative error for a fixed simulation time (that is, the relative efficiency) is bounded as $\varepsilon \to 0$, in agreement with the theory. The coverage function for the different values of ε is displayed Figure 3. It closely follows a uniform distribution as expected whatever the value of ε, showing the robustness of the method.

6.2 Numerical illustration on a larger problem

Consider now a larger example to see how the same simulation method behaves on a large state space. We consider the dodecahedron topology displayed in Figure 4, and the estimation of unreliability between nodes s and t. In this case, deriving analytical results is cumbersome.

The relative efficiency of the method with respect to crude Monte Carlo simulation is displayed Table 3. The degree of improvement as the reliability increases is clearly established. Comparisons with respect to other rare event simulation methods can be found in (Khadiri and Rubino 1996).

Table 2: Results on the simple topology, where the number of replications $n(T)$ corresponds to the *average* number of replications for a simulation time $T = 10$ seconds. This leads to $n(T) = 9e + 06$ for $\varepsilon = 0.5$, $n(T) =$ for $\varepsilon = 0.9$, $n(T) = 7e + 08$ for $\varepsilon = 0.95$, $n(T) = 1.75e + 10$ for $\varepsilon = 0.99$, $n(T) = 6.9e + 10$ for $\varepsilon = 0.995$, $n(T) = 1.71e + 12$ for $\varepsilon = 0.999$ and $n(T) = 1.69e + 14$ for $\varepsilon = 0.9999$.

$r_i\ \forall i$	Est.	Conf. interval	RErr
0.5	3.750e-01	(3.747e-01,3.753e-01)	8.435e-04
0.9	1.901e-02	(1.899e-02,1.903e-02)	1.049e-03
0.95	4.872e-03	(4.866e-03,4.877e-03)	1.059e-03
0.99	1.991e-04	1.988e-04,1.992e-04)	1.050e-03
0.995	4.987e-05	(4.981e-05,4.992e-05)	1.057e-03
0.999	2.000e-06	(1.998e-06,2.002e-06)	1.060e-03
0.9999	2.001e-08	(1.999e-08,2.004e-08)	1.066e-03

Table 3: Results with respect to crude Monte Carlo for the dodecahedron topology. The curves for $\varepsilon = 0.999$ and $\varepsilon = 0.9999$ are not displayed since all the all mass of the empirical distributions is at 1.

$r_i\ \forall i$	Speedup w.r.t. crude MC
0.9	18.9
0.95	188.3
0.98	3800.2

Look now again at the numerical values that can be obtained in practice. We also first consider the case where the number of replications is fixed to $n = 10^4$ and $\varepsilon \to 0$. The first columns of Table 4 display the estimated value, the confidence interval (at confidence level 95%) and the observed Relative Error. It can be immediately seen that, for a fixed number of iterations, the quality of the method is getting worse, until the rare event is not observed anymore (therefore exhibiting the same kind of behavior than crude Monte Carlo would). The last column displays the Kolmogorov-Smirnov statistic (the supremum over the coverage error) when using 500 estimations for the empirical coverage function. The coverage function for the different values of ε is displayed Figure 5. Here again, the degradation as $\varepsilon \to 0$ can be observed.

In Table 5 the same kind of results are displayed, but using the average number of replications for a simulation time fixed to $T = 10$ seconds on our computer, leading to different values of $n(T)$. It can be observed that the method is still very efficient, though it seems that the observed relative error increases a little when ε. Indeed, for this example, BREff has not been proved. All the same, the method is very efficient for highly reliable

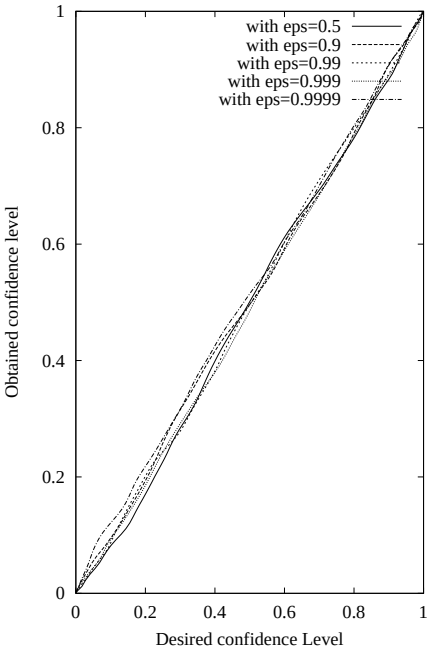

Figure 3: Coverage function for different values of ε and the simple topology for a simulation time fixed to $T = 10$ seconds.

components. The coverage function for the different values of ε is displayed Figure 6. The efficiency is again observed here.

7 CONCLUSIONS

The standard measure of robustness of a rare event estimator in the literature is Bounded Relative Error, stating that the relative width of the confidence interval remains bounded as rarity increases. We show here that this measure is not sufficient since there exist efficient estimators for which the variance is similar to that of crude Monte Carlo (resulting in unbounded relative error), but for which simulation time per run drastically decreases as rarity increases.

In this paper we have proposed the notion of Bounded Relative Efficiency which incorporates both the variance and the computational time. We have also proposed the notion of generalized bounded normal approximation (GBNA) which ensures that the discrepancy between the empirical distribution and the Gaussian one is kept bounded as rarity increases, bounding as a consequence the coverage error of the confidence interval. Since GBNA is only a sufficient condition (and not a necessary one) for ensuring the coverage of the confidence interval, we have also proposed to study the coverage function as rarity increases. All these notions have been illustrated by a problem of estimating the reliability of a static network.

Table 4: Results on the dodecahedron topology, with a number of replications fixed to $n = 10^4$

r_i $\forall i$	Est.	Conf. interval	RErr	KS stat.
0.5	7.082e-01	(6.993e-01,7.171e-01)	1.259e-02	3.3896e-02
0.9	3.200e-03	(2.093e-03,4.307e-03)	3.459e-01	9.6531e-02
0.98	0	(0, 0)	—	7.99e-01
0.99	0	(0, 0)	—	8.28e-01
0.995	0	(0, 0)	—	1
0.999	0	(0, 0)	—	1
0.9999	0	(0, 0)	—	1

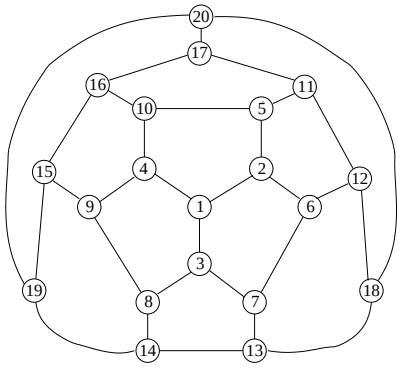

Figure 4: Dodecahedron topology, with $s = 1$ and $t = 20$.

Table 5: Results on the dodecahedron topology, where the number of replications $n(T)$ corresponds to the *average* number of replications for a simulation time $T = 5$ seconds. This leads to $n(T) = 9.8e + 04$ for $\varepsilon = 0.5$, $n(T) = 8.86e + 05$ for $\varepsilon = 0.9$, $n(T) = 6.1e + 07$ for $\varepsilon = 0.98$, $n(T) = 4.55e + 08$ for $\varepsilon = 0.99$, $n(T) = 3.5e + 09$ for $\varepsilon = 0.995$, $n(T) = 4.28e + 11$ for $\varepsilon = 0.999$ and $n(T) = 4.25e + 14$ for $\varepsilon = 0.9999$.

r_i $\forall i$	Est.	Conf. interval	RErr	KS stat.
0.5	7.120e-01	(7.091e-01 , 7.148e-01)	3.98e-03	4.313e-02
0.9	2.889e-03	(2.778e-03 , 3.001e-03)	3.87e-02	7.068e-02
0.98	1.749e-05	(1.644e-05 , 1.854e-05)	6.00e-02	3.036e-02
0.99	2.053e-06	(1.921e-06 , 2.184e-06)	6.41e-02	3.347e-02
0.995	2.540e-07	(2.373e-07 , 2.707e-07)	6.57e-02	2.965e-02
0.999	2.005e-09	(1.870e-09 , 2.1389e-09)	6.69e-02	6.562e-02
0.9999	1.981e-12	(1.847e-12 , 2.115e-12)	6.75e-02	5.436e-02

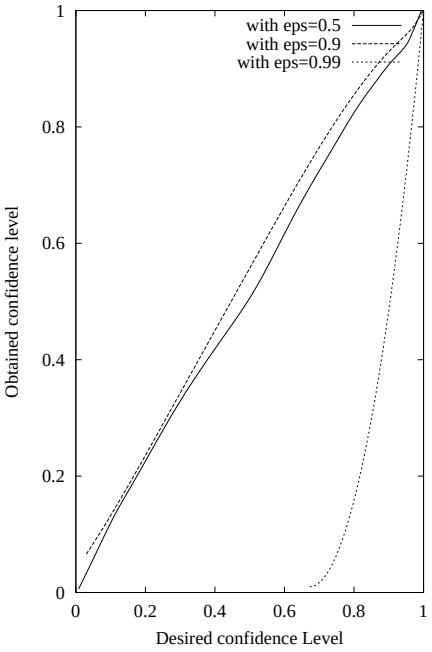

Figure 5: Coverage function for different values of ε and the dodecahedron topology.

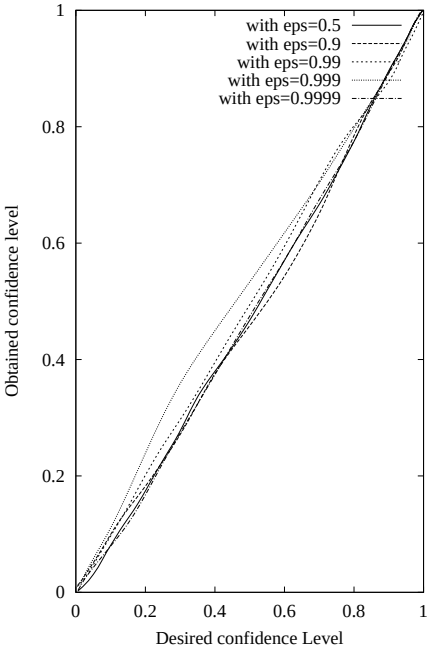

Figure 6: Coverage function for different values of ε and the dodecahedron topology for a simulation time fixed to $T = 10$ seconds.

As directions for future research, we aim at defining necessary and sufficient conditions for obtaining BREff and GBNA in specific contexts such as, for instance, the static reliability problems we have used here. Investigating the importance of other parameters than rarity (e.g. model's size...) is also of interest.

ACKNOWLEDGEMENT

This work has been partially sponsored by EuroNGI Network of Excellence, SurePaths ACI "Sécurité" project, PAIR INRIA associated team and ECOS SUD Project U03E02.

REFERENCES

Fishman, G. 1997. *Monte carlo: Concepts, algorithms and applications*. Springer-Verlag.

Heegaard, P. 1998, May. *Efficient simulation of network performance by importance sampling*. Ph. D. thesis, Norwegian University of Science and Technology.

Heidelberger, P. 1995, January. Fast Simulation of Rare Events in Queueing and Reliability Models. *ACM Transactions on Modeling and Computer Simulation* 5 (1): 43–85.

Heidelberger, P., P. Shahabuddin, and V. Nicola. 1994, April. Bounded Relative Error in estimating transient Measures of Highly Dependable Non-Markovian Systems. *ACM Transactions on Modeling and Computer Simulation* 4 (2): 137–164.

Juneja, S. 1993. *Efficient rare event simulation of stochastic systems*. Ph. D. thesis, Stanford University.

Khadiri, M. E., and G. Rubino. 1996. An efficient formulation of the standard Monte Carlo simulation of binary systems reliability. In *2nd International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*.

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

Rubino, G. 1998. Network reliability evaluation. In *State-of-the art in performance modeling and simulation*, ed. K. Bagchi and J. Walrand. Gordon and Breach Books.

Sandmann, W. 2004. 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*, 49–57: The Operational Research Society.

Schruben, L. 1980. A coverage function for interval estimators of simulation response. *Management Science* 26 (1): 18–27.

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

Tuffin, B. 1999. Bounded Normal Approximation in Simulations of Highly Reliable Markovian Systems. *Journal of Applied Probability* 36 (4): 974–986.

Tuffin, B. 2004, Sept. On numerical problems in simulations of Highly Reliable Markovian Systems. In *1st International Conference on Quantitative Evaluation of SysTems (QEST)*. University of Twente, Enschede, the Netherlands: IEEE CS Press.

AUTHOR BIOGRAPHIES

HECTOR CANCELA is professor and head of the Operations Research Department at the Universidad de la República, Uruguay. His research interests lay mainly in Monte Carlo simulation and combinatorial optimization techniques, applied to the evaluation and design of communication systems. He currently serves in the Board of Directors of the ALIO, the Association of Latin American Operations Research Societies. He is a member of AMS, SIAM and SMAI and of the IFIP WG 7.3. His e-mail address is <cancela@fing.edu.uy>, and his web page is <www.fing.edu.uy/∼cancela>.

GERARDO RUBINO is a senior researcher at INRIA (the French National Institute for Research in Computer Science and Control) where he is the leader of the ARMOR (Architectures and Models of Networks) team at INRIA Unit of Rennes. His research interests are in the quantitative analysis of computer and communication systems, mainly using probabilistic models. He also works on the quantitative evaluation of perceptual quality of multimedia communications over the Internet. He is a member of the IFIP WG 7.3. His e-mail address is <rubino@irisa.fr>, and his Web page is <www.irisa.fr/armor/lesmembres/Rubino/Rubino_en.htm>.

BRUNO TUFFIN received his PhD degree in applied mathematics from the University of Rennes 1 (France) in 1997. Since, he has been with INRIA (Institut National de Recherche en Informatique et Automatique) at Rennes, France. He also spent 8 months at Duke University in 1999. His research interests include developing Monte Carlo and quasi-Monte Carlo simulation techniques for the performance evaluation of computer and telecommunication systems, and more recently developing Internet new pricing schemes. His e-mail address is <btuffin@irisa.fr>, and his web page is <www.irisa.fr/armor/lesmembres/Tuffin/Tuffin_en.htm>.