Bounded Relative Efficiency in Rare Event Simulation

H. Cancela
Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay
cancela@fing.edu.uy

G. Rubino and B. Tuffin
IRISA-INRIA
Campus Universitaire de Beaulieu
35042 Rennes Cedex, France
{rubino,btuffin}@irisa.fr

Abstract

This paper investigates the robustness of rare event simulation estimators when rarity increases. The literature had up to now focused on bounded relative error (BRErr) or bounded normal approximation properties stating respectively that the relative size and the coverage error of the confidence interval are bounded whatever the rarity is. Using a reliability estimation problem, we show that there exists some efficient estimators for which BRErr is not verified. The efficiency is due to the fact that the number of estimations during a given simulation time increases with the rarity. We thus define a property called bounded relative efficiency ecompassing the examples of estimators verifying BRErr, and representing the actual property an analyst has to look at.

1. Introduction

Stochastic models provide a powerful tool to capture the dynamics and behavior of many real-life systems, and as such they are widely and with increasing frequency being used to represent and evaluate different applications. Among the most common, we can mention communication networks and computing systems, but there are many other examples, such as spacecraft and aircraft equipment, industrial production systems, and even social network models, wood-fire propagation models, etc. The use of Monte Carlo simulation is a method of choice for evaluating stochastic models, specially useful when the model size is large and when there are no underlying symmetries that can lead to efficient use of other analytical or numerical techniques.

In many situations, the properties of interest depend critically on the occurrence of a rare event, that is, on observing a subset of system states which appear with very low probability. For instance, if we consider a communication network, we can be interested in the probability of packet loss due to buffer overfilling, or in the probability of not being able to connect two terminal sites of the network, due to failures in the links; usually these events have very small probabilities (maybe smaller than 10^{-9} in actual practice).

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. 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 methods, offering different tradeoffs, the immediate question is how to choose the most appropriate one. 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, and bounded normal approximation, discussed in more detail in Section 2. These

concepts focus on the precision attained and the robustness of the simulation estimators (both are important features, see for example the discussion in [2]), but do not take into account the computational times associated with them.

In this work, we combine some features of the two approaches: we focus on the overall efficiency of simulation methods (combining in a single measure the precision attained and the computational times employed), but we follow the analytical approach. In particular, in Section 3 we 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. In Section 4, we analyze a particular simulation method, and we observe that it fulfills a sufficient condition for Bounded Relative Efficiency to hold, implying its robust behavior.

2. Properties of rare event estimators

We consider the estimation of a rare event with probability γ. We characterize the rare event by a rarity parameter ε so that as $\varepsilon \to 0$, $\gamma \to 0$. In reliability models for instance, ε may represent a maximum failure rate or the reliability of a component [8]. In queuing models, it may be $\varepsilon = 1/B$ where B is the buffer size, so that the buffer overflow probability $\gamma \to 0$ as $\varepsilon \to 0$ [2].

2.1. Bounded Relative Error

Let us consider an unbiased estimator $\hat{\gamma}$ of γ taken from a sample having size n. Bounded Relative Error (BRErr) has been defined in [8], and further studied in [3, 4, 5], 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. Formally:

Definition 1 *Let σ_n^2 denote the variance of $\hat{\gamma}$ for a sample size n and be z_δ the $1 - \delta/2$ quantile of the standard normal distribution. Then the relative error RErr is defined by*

$$RErr = z_\delta \frac{\sqrt{\sigma_n^2}}{\gamma}. \qquad (1)$$

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

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: 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^*.

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

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

Note that this quantity is always less than or equal to 2.

In [7], the relations between asymptotic optimality and BRErr are investigated.

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) [9] 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 its variance), $\mathcal{N}$ the standard normal distribution, $\hat{\gamma} = n^{-1} \sum_{i=1}^n X_i$, $\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^I (X_i - \hat{\gamma})^2$ and F_n the distribution of the centered and normalized sum $(\hat{\gamma} - \gamma)/\hat{\sigma}_n$, then there exists an absolute constant $a > 0$ such that, for each x and I

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

Definition 3 *We say that $\hat{\gamma}$ 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 [10], it is shown that, for Markovian reliability models, BNA implies that the estimation of the variance is asymptotically correct, implying 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 [10].

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 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 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 [6]). 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 has also cost $\Theta(M)$. The usual situation is that the reliabilities of the lines are high, leading to a rare event situation ($1 - R \approx 0$).

In [1], a different and simple 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 average, we have to wait for $\mathrm{E}(F) = 1/q$ samples of G' to generate 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. The idea is then to sample first 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 lot of computations as the reliability increases). It remains to know if they are connected in the fth or not. Then we must build that fth copy, sampling 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 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

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

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 up, link i_{h,w_h} is 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 [1], 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 $\gamma = 1 - R \rightarrow 0$ as $\varepsilon \rightarrow 0$. Let $\hat{\gamma}$ be the above estimator of the unreliability.

The Relative Error of this method is

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

as $\varepsilon \rightarrow 0$. 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 $\hat{\gamma}$ be an estimator of γ, and σ_n^2 be its variance when using n replications (possibly dependent). Let t_n be the average simulation time to get those n replications. The relative efficiency of $\hat{\gamma}$ is given by*

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

We will say that $\hat{\gamma}$ 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).

4. 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.

Let $r > 0$ be the real such that $\gamma = \Theta(\varepsilon^r)$ and 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') with $r' > r$ (see [6] for general discussions on paths and cuts).

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

$$q = \prod_{h=1}^{H}(1-p_h) = \prod_{h=1}^{H}\left(1 - \prod_{k=1}^{K_h}\left(1 - a_{i_{h,k}}\varepsilon^{b_{i_{h,k}}}\right)\right)$$

$$= \prod_{h=1}^{H}\Theta(a_h\varepsilon^{b_h}) = (\prod_{h=1}^{H}a_h)\Theta(\varepsilon^{\sum_{h=1}^{H}b_h}).$$

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. Numerical examples are not provided due to lack of room.

Acknowledgement

References

[1] M. El Khadiri and G. Rubino. 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*, 1996.

[2] P. Heegaard. *Efficient simulation of network performance by Importance Sampling*. PhD thesis, Norwegian University of Science and Technology, May 1998.

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

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

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

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

[7] 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*, pages 49–57, 2004.

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

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

[10] B. Tuffin. 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, Sept 2004. IEEE CS Press.