

An Adaptive Zero-Variance Importance Sampling Approximation for Static Network Dependability Evaluation

Bruno Tuffin · Samira Saggadi · Pierre L'Ecuyer

Received: date / Accepted: date

Abstract We propose an adaptive parameterized method to approximate the zero-variance change of measure for the evaluation of static network reliability models, with links subject to failures. The method uses two rough approximations of the unreliability function, conditional on the states of any subset of links being fixed. One of these approximation, based on mincuts, under-estimates the true unknown unreliability, whereas the other one, based on minpaths, over-estimates it. Our proposed change of measure takes a convex linear combination of the two, estimates the optimal (graph-dependent) coefficient in this combination from pilot runs, and uses the resulting conditional unreliability approximation at each step of a dynamic importance sampling algorithm. This new scheme is more general and more flexible than a previously-proposed zero-variance approximation scheme, based on mincuts only, and which was shown to be robust asymptotically when unreliabilities of individual links decrease toward zero. Our numerical examples show that the new scheme is often more efficient when the unreliabilities of the individual links are not so small but the overall unreliability is small because the system can fail in many ways.

1 Introduction

This paper investigates the evaluation of the probability that a given set of nodes are not connected, in a network represented by a graph where links are considered failed with specified probabilities. This type of problem finds applications in areas such as

B. Tuffin and S. Saggadi
INRIA Rennes – Bretagne Atlantique, Campus Universitaire de Beaulieu, 35042 Rennes Cedex, FRANCE
Tel.: +33-2-99-84-74-94
E-mail: {bruno.tuffin,samira.saggadi}@inria.fr

P. L'Ecuyer
Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, C.P. 6128, Succ.
Centre-Ville, Montréal, H3C 3J7, Canada.
E-mail: lecuyer@iro.umontreal.ca

telecommunications, power supply systems, and transportation systems, among others [3–5]. In practical situations, one typically faces large topologies and estimations of very small probabilities. When dealing with large networks, Monte Carlo simulation techniques are generally needed, because the exact evaluation is an NP-hard problem in general [2], and variance reduction methods are required for an efficient estimation of rare events [19]. The present work extends that of [15], where an importance sampling (IS) technique approximating the zero-variance estimator was developed. The approximation in [15] is based on the computation of probabilities of minimal cuts in subgraphs. The authors prove that in an asymptotic rare-event regime where the failure probabilities of individual links become very small, the RE of their IS estimator remains bounded, and can even converge to zero under additional conditions, a property rarely verified by other rare event estimators. For general background on IS and zero-variance approximation, we refer the reader to [1, 6, 7, 14, 12, 15], and references therein.

On the other hand, there are situations where this assumption that failure probabilities of individual links are very small is not realistic and is not a requirement for the set of considered nodes not being connected to be a rare event. The probability of this event can be very small for example because there is a large number of possible paths connecting the considered nodes, and this may happen even when the links have a large failure probability. In this case, the unreliability approximation based on mincuts may underestimate the true unreliability by a large factor. The goal and contribution of this paper is to illustrate how the zero-variance approximation based on mincuts can deteriorate in this type of situation, and to propose an alternative (more general) estimator. We first introduce a second approximation, based on the probabilities of minpaths, which overestimates the true value. Then we use a convex linear combination of the two approximations to estimate the conditional unreliability of subgraphs encountered during the sequential IS procedure. The coefficient in this linear combination is learnt (or approximately optimized) via pilot simulation runs in a preliminary phase. With the optimal coefficient, the variance of the resulting estimator is necessarily smaller than the one in [15], which is a particular case, and is robust in more situations where system failure is a rare event. Note that this new estimator is not equivalent to taking a linear combination of two IS estimators, which would amount to using the difference between the two as a control variate.

The rest of the paper is organized as follows. In Section 2, we present the model and recall the zero-variance IS scheme and its mincut-maxprob approximation proposed earlier. In Section 3, we start with a motivating example where the mincut-maxprob approximation does not work well, introduce a new zero-variance approximation based on minpaths, and finally propose an IS estimator based on the convex linear combination of the two approximations at each step. In Section 4, we show how the optimal coefficient in this linear combination can be estimated. In Section 5, we illustrate the benefits of our method via several examples. Conclusions and directions for future research are given in Section 6.

2 Model and simulation by approximate zero-variance importance sampling

We consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{K})$ where \mathcal{V} is the set of nodes, $\mathcal{E} = \{1, \dots, m\}$ is the set of links connecting nodes, and \mathcal{K} is a subset of the node-set, called the *terminal set*. Nodes are assumed perfect, i.e., they do not fail, while links can fail, link $i \in \mathcal{E}$ failing with probability q_i . All failure events of individual links are assumed independent. Our goal is to compute the probability $u = u(\mathcal{G})$ that the nodes in the terminal set \mathcal{K} are not all connected by working links (i.e., not in the same connected component), in which case we say that the system is failed. The random configuration of the graph is fully characterized by vector

$$X = (X_1, \dots, X_m),$$

where the X_i , for $1 \leq i \leq m$, are independent Bernoulli random variables with parameters $1 - q_i$, giving $X_i = 1$ if link i is working, and 0 if it is failed,

We define the function $\psi : \{0, 1\}^m \rightarrow \{0, 1\}$ by $\psi(x) = 1$ if the nodes in \mathcal{K} are not connected when the graph configuration is $x \in \{0, 1\}^m$, and 0 otherwise. We then have

$$u = u(\mathcal{G}) = \mathbb{E}[\psi(X)] = \sum_{x \in \{0, 1\}^m} \psi(x) \prod_{i=1}^m ((1 - x_i)q_i + x_i(1 - q_i)).$$

Computing this value is known to be NP-hard in general [2], because the number of graph configurations grows exponentially with the number m of links, and Monte Carlo simulation becomes a relevant tool. Though, standard Monte Carlo, consisting in approximating u by the average over independent realizations of $\psi(X)$ is inefficient because the failure event for the system happens too rarely and may even never be observed. Specific rare-event simulation techniques have to be developed.

We refer the reader to [19] for general descriptions of rare-event simulation techniques and applications, and to [4] for a state-of-the-art review specific to static dependability analysis. Here we build on the zero-variance sequential IS formalism introduced in [15], defined as follows. The link states X_1, \dots, X_m are generated successively, in that order, under an IS scheme that replaces the original probability $q_i = \mathbb{P}[X_i = 0]$ to generate X_i by a new probability \tilde{q}_i , which may depend on the previously generated link states X_1, \dots, X_{i-1} . Denoting

$$L(x) = \prod_{i=1}^m \left(x_i \frac{1 - q_i}{1 - \tilde{q}_i} + (1 - x_i) \frac{q_i}{\tilde{q}_i} \right),$$

the *likelihood ratio* of original and new probabilities, we have that $\psi(X)L(X)$ is an unbiased estimator of u . Moreover, if we define

$$u_i(x_1, \dots, x_{i-1}) = \mathbb{E}[\psi(X) | X_1 = x_1, \dots, X_{i-1} = x_{i-1}],$$

the unreliability of graph \mathcal{G} conditional on the states of links 1 to $i - 1$, and if we put $\tilde{q}_i = \tilde{q}_i^*$ where

$$\begin{aligned} \tilde{q}_i^* &\stackrel{\text{def}}{=} \tilde{\mathbb{P}}[X_i = 0 \mid X_1 = x_1, \dots, X_{i-1} = x_{i-1}] \\ &= \frac{q_i u_{i+1}(x_1, \dots, x_{i-1}, 0)}{q_i u_{i+1}(x_1, \dots, x_{i-1}, 0) + (1 - q_i) u_{i+1}(x_1, \dots, x_{i-1}, 1)} \end{aligned} \quad (1)$$

$$= q_i \frac{u_{i+1}(x_1, \dots, x_{i-1}, 0)}{u_i(x_1, \dots, x_{i-1})}, \quad (2)$$

then the unbiased estimator $\psi(X)L(X)$ has zero variance (that is, it always provides the exact value); see [15] for the details and proofs.

However, implementing this estimator would require the knowledge of all the $u_i(x_1, \dots, x_{i-1})$, and if we knew them, there would be no need for simulation. The principle of zero-variance approximation is to use simple approximations $\hat{u}_{i+1}(\cdot)$ of $u_{i+1}(\cdot)$ in (1), so that (1) is replaced by

$$\begin{aligned} \tilde{q}_i &= \tilde{\mathbb{P}}[X_i = 0 \mid X_1 = x_1, \dots, X_{i-1} = x_{i-1}] \\ &= \frac{q_i \hat{u}_{i+1}(x_1, \dots, x_{i-1}, 0)}{q_i \hat{u}_{i+1}(x_1, \dots, x_{i-1}, 0) + (1 - q_i) \hat{u}_{i+1}(x_1, \dots, x_{i-1}, 1)}. \end{aligned} \quad (3)$$

This way, if the approximation $\hat{u}_{i+1}(\cdot)$ is good enough, the sampling probabilities will be close to the optimal ones and we can expect the variance to be reduced by large factors. It is proposed in [15] to use for $\hat{u}_{i+1}(\cdot)$ the maximal probability of a mincut of the subgraph obtained when x_1, \dots, x_i are fixed. More specifically, a *mincut* in \mathcal{G} is a subset \mathcal{C} of links such that when all the links in \mathcal{C} are failed, then the nodes of \mathcal{K} cannot be all connected (the system is necessarily failed), and no strict subset of \mathcal{C} has this property. A mincut with maximal probability is one for which the probability that all its links are failed is maximal among all mincuts. This maximal probability is called the *mincut-maxprob* approximation, and can be computed in polynomial time in m . When applying IS with this zero-variance approximation, we take $\hat{u}_{i+1}(x_1, \dots, x_i)$ in (3) as the probability of a mincut with maximal probability in the subgraph where (X_1, \dots, X_i) are fixed to (x_1, \dots, x_i) (the links $e \leq i$ with $X_e = 0$ are removed from the graph and those with $X_e = 1$ cannot fail, so we can merge the nodes that they connect). We shall denote the resulting $\hat{u}_{i+1}(\cdot)$ by $\hat{u}_{i+1}^{\text{mc}}(\cdot)$. It is shown in [15] that this approximation leads to an asymptotically robust estimator when $q_i \rightarrow 0 \forall i \in \mathcal{E}$, in the following sense.

Recall that for an unbiased estimator Y of an unknown quantity u , the variance is $\text{Var}[Y] = \mathbb{E}[(Y - u)^2]$, the relative error (RE) is $\text{RE}[Y] = (\text{Var}[Y])^{1/2}/u$, and the *relative centered moment* of order β is $\mathbb{E}[(Y - u)/u]^\beta$, for any $\beta \geq 1$. If Y depends on a parameter ε and if we consider an asymptotic regime under which $\varepsilon \rightarrow 0$, then $Y = Y(\varepsilon)$ is said to have bounded [vanishing] relative centered moment of order β in this asymptotic regime if $\mathbb{E}[(Y(\varepsilon) - u(\varepsilon))/u(\varepsilon)]^\beta$ remains bounded [converges to 0] when $\varepsilon \rightarrow 0$ [10]. When this holds for $\beta = 2$, that is, if $\text{RE}[Y(\varepsilon)]$ remains bounded [converges to 0] when $\varepsilon \rightarrow 0$, the estimator is said to have *bounded [vanishing] RE*. It is proved in [15] that the dynamic IS estimator based on (3) with the mincut-maxprob approximation always has bounded RE, and has vanishing RE under some additional

conditions, in an asymptotic regime where $\max_i q_i \rightarrow 0$ when $\varepsilon \rightarrow 0$. These results can be extended to bounded and vanishing relative moments of order $\beta \geq 1$.

3 A minpath-maxprob approximation and a linear combination of two approximations

In this section, we exhibit examples where the mincut-maxprob approximation does not work well, we propose the minpath-maxprob approximation as an alternative, then we propose a convex linear combination of the two approximations, which includes each of the two as a special case and is more general.

3.1 A motivating example

Example 1 Consider the graph topology described in Figure 1, comprised of a grid of r rows and two columns for the intermediate nodes. We look for the probability that the grey nodes s and t are not connected given that all links $i \in \mathcal{E}$ have a probability $q_i = q$ to fail. Note that in the original graph, there are three (vertical) mincuts with maximal probability q^r , so $\hat{u}_1^{\text{mc}}(\emptyset) = q^r$. There are also several mincuts of probabilities q^{r+1} , q^{r+2} , etc.

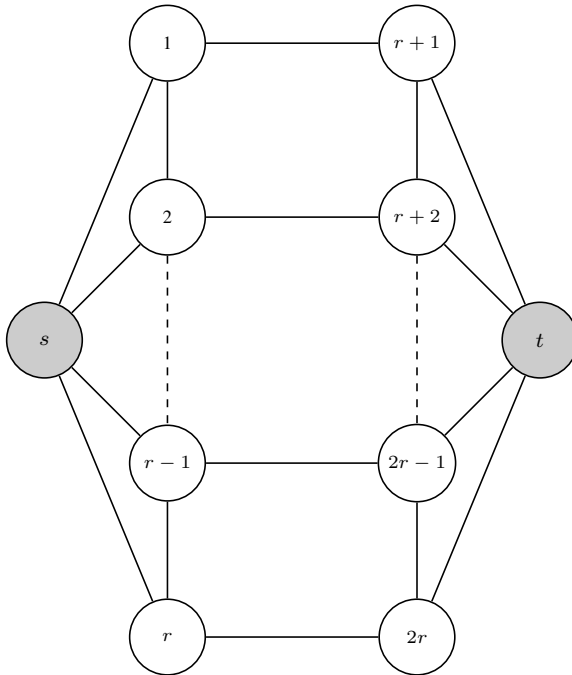


Fig. 1 A topology made of r rows of two links each.

We estimated the unreliability of this network by simulation with the mincut-maxprob approximation, for various values of r , while adjusting the unreliabilities $q_i = q$ of the individual links deterministically in a way that the overall unreliability u was near 10^{-8} for all r . The sample size was $n = 10^5$ in all cases. Table 1 reports the empirical unreliability \hat{u} and the estimated RE (which we adopt here as a measure of effectiveness of the estimator) denoted $\widehat{\text{RE}}$, as functions of r . The (estimated) RE tends to increase with r , although there is noise in its estimation. For $r = 200$ (last row in the table), the unreliability and the RE turn out to be highly underestimated, and this happens because the most relevant configurations are rare and not sampled anymore (or sufficiently), leading to misleading results. When this happens, a confidence interval on u computed in a standard way is unlikely to cover the true value of u . The last column gives the mincut-maxprob approximation for the entire graph, $\hat{u}_1^{\text{mc}}(\emptyset)$. This approximation of u becomes much too small when r increases.

Table 1 Empirical mean \hat{u} and $\widehat{\text{RE}}$ of the IS estimator of u based on the mincut-maxprob approximation with sample size $n = 10^5$, for the graph of Figure 1, for selected values of r , and q chosen so that the estimated u is near 10^{-8} . The last column gives the mincut-maxprob approximation of u for the entire graph.

r	q	$10^8 \hat{u}$	$\widehat{\text{RE}}$	$\hat{u}_1^{\text{mc}}(\emptyset) = q^r$
2	0.00007	1.46	0.33	4.9×10^{-9}
5	0.02	1.06	0.46	3.2×10^{-9}
10	0.1245	1.11	1.8	8.9×10^{-10}
30	0.371	1.14	7.9	1.2×10^{-13}
40	0.427	1.05	9.9	1.6×10^{-15}
50	0.4665	1.08	31	2.7×10^{-17}
70	0.521	1.35	22	1.5×10^{-20}
100	0.575	1.48	40	9.2×10^{-25}
200	0.655	0.48	44	1.8×10^{-37}

The poor behavior of the estimator for this example can be explained as follows. The mincut-maxprob approximation was proved in [15] to provide a bounded RE for the unreliability when $q \rightarrow 0$, using the fact that $u = u(\mathcal{G})$ and the maximum probability of a mincut are then of the same order of magnitude. But here, q increases with r , and the probability of a mincut is not a good approximation of u because there are often too many mincuts with comparable probabilities. As an illustration, for the entire graph \mathcal{G} , the mincut-maxprob approximation gives $\hat{q}(\mathcal{G}) = \hat{u}_1^{\text{mc}}(\emptyset) = q^r$, because the smallest mincuts are of cardinality r . But by comparing q^r to the unreliability estimate \hat{u} given in Table 1 (which is near 10^{-8}), in the fifth and third columns, we find huge gaps when r is large. For $r = 100$, for example, $q^r = 0.575^{100} \approx 9.2 \times 10^{-25}$. In other words, here the asymptotic regime where $q \rightarrow 0$ is not the appropriate one when r is large and q is not very small.

3.2 The minpath-maxprob approximation

Define a *path* (or \mathcal{K} -path) in the graph \mathcal{G} as a set P of links i such that when these links are up ($X_i = 1$ for all $i \in P$), the nodes in \mathcal{K} are in the same connected component of the resulting graph, whatever the state of other links. A *minpath* of \mathcal{G} is a path $P = P(\mathcal{G})$ for which no strict subset of $P(\mathcal{G})$ is also a path. Let $\mathcal{F}_{\mathcal{G}}$ be the set of minpaths of \mathcal{G} . The probability that a minpath P has all its links up is $p = p(P) = \prod_{i \in P(\mathcal{G})} (1 - q_i)$. The *minpath-maxprob* approximation of the unreliability of graph \mathcal{G} is then

$$\hat{u}^{\text{mp}}(\mathcal{G}) = 1 - \max_{P \in \mathcal{F}_{\mathcal{G}}} p(P).$$

This approximation is the largest probability among all minpaths.

To implement IS with the minpath-maxprob approximation, we proceed in the same way as with the mincut-maxprob approximation, except that the approximations $\hat{u}_{i+1}^{\text{mc}}(\cdot)$ used at each step are replaced by $\hat{u}_{i+1}^{\text{mp}}(\cdot)$, defined as follows. At each step $i+1$, for $i = 0, \dots, i-1$, the states of the first i links have been sampled already and fixed to x_1, \dots, x_i , respectively. We define the graph $\mathcal{G}_i = \mathcal{G}_i(x_1, \dots, x_i)$ obtained from \mathcal{G} by removing all links $j \leq i$ for which $x_j = 0$ and forcing the links j such that $x_j = 1$ to be operational. We then define $\hat{u}_{i+1}^{\text{mp}}(x_1, \dots, x_i)$ as the *minpath-maxprob* approximation of graph \mathcal{G}_i , but not considering in the computations the probabilities of links $j \leq i$ with $x_j = 1$, because those links are known to be operational. This is similar to the definition of $\hat{u}_{i+1}^{\text{mc}}(x_1, \dots, x_i)$ based on mincuts.

For any $P' \in \mathcal{F}_{\mathcal{G}}$, let $E(P')$ be the event that “all links of P' are working.” Then, for any $P \in \mathcal{F}_{\mathcal{G}}$,

$$u = 1 - \mathbb{P}[\cup_{P' \in \mathcal{F}_{\mathcal{G}}} E(P')] \leq 1 - \mathbb{P}[E(P)] \leq \hat{u}^{\text{mp}}(\mathcal{G}).$$

The minpath-maxprob approximation therefore over-estimates the true conditional unreliability, as opposed to the mincut-maxprob approximation which under-estimates it.

It was shown in [15] that for a fixed topology, IS with the mincut-maxprob approximation provides an estimator of the unreliability with bounded RE when $q_i \rightarrow 0 \forall i$, essentially because $\hat{u}_{i+1}^{\text{mc}}(\cdot) = \Theta(u_{i+1}(\cdot)) \forall i$. Using similar arguments, one can show that IS with the minpath-maxprob approximation provides an estimator with bounded RE for the *reliability* $1 - u$ when $q_i \rightarrow 1 \forall i$, because $1 - \hat{u}_{i+1}^{\text{mp}}(\cdot) = \Theta(1 - u_{i+1}(\cdot)) \forall i$.

Example 1 (Continued). We applied IS to the same example as in Section 3.1, with the same experimental setting and sample size, but this time with the minpath-maxprob approximation. The results are given in Table 2. We observe (empirically) that this time, the RE *decreases* with r , and that the method does better than the mincut-maxprob approximation when r is large. Here, in the original graph, there are r (horizontal) minpaths with maximal probability $p = (1 - q)^3$, so the minpath-maxprob approximation for the original graph is $\hat{u}_1^{\text{mp}}(\emptyset) = 1 - (1 - q)^3$. It is given in the last column of the table. There are also several minpaths of probabilities $(1 - q)^4$, $(1 - q)^5$, etc. We see that the approximation $\hat{u}_1^{\text{mp}}(\emptyset)$ overestimates u by a factor of approximately 10^8 when r is large, but it nevertheless provides an IS estimate with much less variance than the one based on $\hat{u}_1^{\text{mc}}(\emptyset)$.

Table 2 Empirical mean \hat{u} and $\widehat{\text{RE}}$ of the IS estimator of u based on the minpath-maxprob approximation with sample size $n = 10^5$, for the graph of Figure 1, for selected values of r , and q chosen so that the estimated u is near 10^{-8} . The last column gives the mincut-maxprob approximation of u for the entire graph.

r	q	$10^8 \hat{u}$	$\widehat{\text{RE}}$	$\hat{u}_1^{\text{mp}}(\emptyset)$
2	0.00007	1.68	66	0.0002
5	0.02	3.18	160	0.058
10	0.1245	1.15	110	0.32
30	0.371	1.36	75	0.75
40	0.427	1.20	36	0.81
50	0.4665	0.98	26	0.84
70	0.521	1.58	17	0.89
90	0.559	1.19	6.6	0.91
100	0.575	1.52	9.8	0.92
200	0.655	1.13	3.9	0.95

3.3 A linear combination of two unreliability approximations

What we now propose is an IS scheme that approximates the unreliability functions $u_{i+1}(\cdot)$ in the zero-variance scheme (3) by a convex linear combination of $\hat{u}_{i+1}^{\text{mc}}$ and $\hat{u}_{i+1}^{\text{mp}}$. That is, we take

$$\hat{u}_{i+1}(x_1, \dots, x_i) = \alpha \hat{u}_{i+1}^{\text{mc}}(x_1, \dots, x_i) + (1 - \alpha) \hat{u}_{i+1}^{\text{mp}}(x_1, \dots, x_i) \quad (4)$$

$\forall i$ and $\forall (x_1, \dots, x_i) \in \{0, 1\}^s$, for some coefficient $\alpha \in [0, 1]$ that remains to be chosen. We know that for each choice of (x_1, \dots, x_i) , there is an $\alpha = \alpha(x_1, \dots, x_i) \in [0, 1]$ such that (4) is exactly equal to $u_{i+1}(x_1, \dots, x_i)$, because the right side is a convex linear combination of one term that underestimates and one term that overestimates. So by choosing a different α for each configuration (x_1, \dots, x_i) , it would be possible in principle to achieve zero variance. But optimizing all these different α 's would be difficult. Here we simplify by imposing a single α for all stages and configurations. Ideally, we want to select α to minimize the variance $V(\alpha)$ of the resulting (unbiased) IS estimator $\psi(X)L(X)$, as a function of α . For that, we will assume that V is differentiable and unimodal, with a minimum in $(0, 1)$. Under this condition, minimizing $V(\alpha)$ is equivalent to finding a root of the equation $V'(\alpha) \stackrel{\text{def}}{=}} (\partial V / \partial \alpha)(\alpha) = 0$. (We do not have a proof that the condition holds in all cases, but this has worked well in all the examples we tried.) Such a root can be approximated in pilot experiments, as explained in the next section. By using a linear combination with a nearly-optimal linear coefficient α , the IS method can self-adapt to the underlying asymptotic regime of the considered graph topology.

Note that instead of using a single real-valued coefficient α for all i , we could take a vector $(\alpha_1, \dots, \alpha_m)$ where α_i is used at step i of the IS algorithm, when sampling the i th link. This increased flexibility could improve the performance of the combined method. On the other hand, optimizing the vector is likely to become more difficult (we would face a high-dimensional optimization problem). Moreover, the best α_i could also depend strongly on (x_1, \dots, x_{i-1}) , so an α_i that depends on i would also be an average over several configurations. In this paper, we decided to avoid these

complications and just take a single constant α . We will see that this simple approach already provides a significant gain in some situations.

4 Learning a good coefficient α for the linear combination

We suggest in this section two ways of approximating the value α^* of α that minimizes the variance of the resulting estimator.

4.1 A heuristic based on the total unreliability estimation

If we knew u , we could compute the value of α for which the linear combination would give the exact value of u when $i = 1$; that is, for which

$$u = u_1(\emptyset) = \alpha \widehat{u}_1^{\text{mc}}(\emptyset) + (1 - \alpha) \widehat{u}_1^{\text{mp}}(\emptyset).$$

This gives

$$\alpha = \frac{\widehat{u}^{\text{mp}}(\emptyset) - u}{\widehat{u}^{\text{mp}}(\emptyset) - \widehat{u}^{\text{mc}}(\emptyset)}, \quad (5)$$

in which the mincut-maxprob and minpath-maxprob approximations $\widehat{u}^{\text{mp}}(\emptyset)$ and $\widehat{u}^{\text{mc}}(\emptyset)$ are easily computed. This α is always in the interval $[0, 1]$. It is generally not optimal, because it gives an exact value only for $i = 1$, but it can be the base of a heuristic. We can replace the unknown u by a rough estimate $\widehat{u}_{n_0}(\mathcal{G})$ obtained from pilot runs, for instance the IS method with mincut-maxprob approximations, using a small to moderate sample size n_0 . This yields

$$\alpha = \alpha_{\text{tot}} \stackrel{\text{def}}{=} \frac{\widehat{u}^{\text{mp}}(\emptyset) - \widehat{u}_{n_0}(\mathcal{G})}{\widehat{u}^{\text{mp}}(\emptyset) - \widehat{u}^{\text{mc}}(\emptyset)}. \quad (6)$$

Because of the noise in $\widehat{u}_{n_0}(\mathcal{G})$, this α_{tot} may end up outside $[0, 1]$, in which case we can project it to the interval $[0, 1]$ (take the nearest boundary).

4.2 A Robbins-Monro algorithm

A second approach to estimate the optimal α is to use an iterative stochastic approximation (SA) method to estimate a root of the equation $V'(\alpha) = 0$. The procedure starts with an arbitrary $\alpha_0 \in (0, 1)$, for example $\alpha_0 = 0.5$, or α_0 given by the heuristic of Section 4.1, or α_0 taken as a crude estimate of α^* if one is available, and iterates as

$$\alpha_{\ell+1} = \alpha_{\ell} - a_{\ell} H_{\ell}^{-1} \widehat{V}'(\alpha_{\ell}), \quad (7)$$

where $\widehat{V}'(\alpha_{\ell})$ is an estimate of $V'(\alpha_{\ell})$ (we explain in the next subsection how it can be obtained), H_{ℓ} is an estimate of $V''(\alpha_{\ell})$ in the case of second order SA algorithm and is just taken as equal to 1 for a first-order algorithm, and the pre-specified sequence of *step sizes* $\{a_{\ell}, \ell \geq 0\}$ is a positive decreasing sequence that satisfies $\sum_{\ell=1}^{\infty} a_{\ell} = \infty$ and $\sum_{\ell=1}^{\infty} a_{\ell}^2 < \infty$, to ensure convergence to the minimum if V is

unimodal and smooth. There is a vast literature on SA methods and their properties; see for example [8, 11, 16, 17, 22, 23] and references therein. We take a_ℓ of the general form

$$a_\ell = \frac{e}{(C + \ell)^\beta},$$

where $e > 0$, $C \geq 0$, and $\beta \in (1/2, 1]$ are tunable constants [17]. If the derivative estimator used at each step is unbiased, if $H_\ell \rightarrow H$ when $\ell \rightarrow \infty$, and under mild additional conditions, the best convergence rate of α_ℓ to α^* when $\ell \rightarrow \infty$ is obtained (only) when $\beta = 1$ and $2eV''(\alpha^*)/H \geq 1$, and then α_ℓ converges to α^* at the canonical rate of $\mathcal{O}(\ell^{-1/2})$. Moreover, the asymptotic variance constant $\lim_{\ell \rightarrow \infty} \ell^{-1} \text{Var}[\alpha_\ell]$ is minimized by taking $a_\ell H_\ell^{-1} = (\ell V''(\alpha^*))^{-1}$. However, finite-sample behavior is often better with a smaller β ; that is, with a step size that decreases more slowly. For example, [17] recommend $\beta = 0.6$. The performance is also sensitive to the choice of e and C [11, 17, 23]. A more robust and stable approach is to retain as the solution the average

$$\bar{\alpha}_{\ell_0, \ell} = \frac{1}{\ell - \ell_0} \sum_{\iota=\ell_0+1}^{\ell} \alpha_\iota, \quad (8)$$

for some fixed positive integer ℓ_0 . This $\bar{\alpha}_{\ell_0, \ell}$ converges to α^* at the canonical rate of $\mathcal{O}(\ell^{-1/2})$ even for $\beta < 1$ and for arbitrary $e > 0$ and $C > 0$ [8, 17, 18].

For our experiments, we used only the first-order algorithm ($H_\ell = 1$) and retained the average (8) as our estimate of the optimal α , with $\ell_0 \approx \ell/2$. We also took $\beta = 0.6$, $C = 1$, e chosen so that $\alpha_1 - \alpha_0 = eV'(\alpha_0) \approx \min(\alpha_0, 1 - \alpha_0)/10$, and $\alpha_0 = \alpha_{\text{tot}}$ from the heuristic in (6). This choice of e requires a preliminary estimation of $V'(\alpha_0)$, which we obtained using 10^3 points. The idea behind this strategy is to take (arbitrarily) $1/10$ of a maximal move of $\min(\alpha_0, 1 - \alpha_0)$, where this move is such that when we are close to 0 or 1, we do not want to move too far, to avoid being trapped in a “bad” area; for example an area where the variance is poorly estimated due to important rare events and/or where the variance is almost constant as a function of α .

4.3 Estimating the Derivative

We now turn to the question of how to estimate the derivative $V'(\alpha_\ell)$ at each step of the SA algorithm. To write an explicit expression for $V(\alpha)$, we introduce some notation. Let p_{i, x_i} denote the probability that the state of link i is x_i , that is, $p_{i,1} = p_i = 1 - q_i$ and $p_{i,0} = q_i$, let

$$\hat{u}_{i, \alpha}(x_i) = \alpha \hat{u}_{i+1}^{\text{mc}}(x_1, \dots, x_i) + (1 - \alpha) \hat{u}_{i+1}^{\text{mp}}(x_1, \dots, x_i)$$

the approximation of u conditional on x_1, \dots, x_{i-1} link i being in state x_i (we abusively omit from the notation the conditioning on the states of the first $i - 1$ links), and let

$$\hat{u}_{i, \alpha}^{\text{tot}} = p_i \hat{u}_{i, \alpha}(1) + q_i \hat{u}_{i, \alpha}(0).$$

Let $\tilde{\mathbb{E}}_\alpha$ and $\tilde{\mathbb{P}}_\alpha$ be the expectation and probability under IS with the linear coefficient set to α . Then we can write

$$\begin{aligned}
V(\alpha) &= \tilde{\mathbb{E}}_\alpha [\psi(X)L^2(X)] \\
&= \sum_{x \in \{0,1\}^m} \psi(x) \frac{(\mathbb{P}[X=x])^2}{\tilde{\mathbb{P}}_\alpha[X=x]} \\
&= \sum_{x \in \{0,1\}^m} \psi(x) \prod_{i=1}^m \frac{p_{i,x_i} \hat{u}_{i,\alpha}^{\text{tot}}}{\hat{u}_{i,\alpha}(x_i)} \\
&= \sum_{x \in \{0,1\}^m} \psi(x) \prod_{i=1}^m \left(p_{i,x_i} \frac{\alpha \hat{v}_i^{\text{tot}} + \hat{u}_i^{\text{mp,tot}}}{\alpha \hat{v}_i(x_i) + \hat{u}_i^{\text{mp}}(x_i)} \right), \tag{9}
\end{aligned}$$

with

$$\begin{aligned}
\hat{v}_i(x_i) &= \hat{u}_{i+1}^{\text{mc}}(x_1, \dots, x_i) - \hat{u}_{i+1}^{\text{mp}}(x_1, \dots, x_i) \\
\hat{u}_i^{\text{mp}}(x_i) &= \hat{u}_{i+1}^{\text{mp}}(x_1, \dots, x_i) \\
\hat{v}_i^{\text{tot}} &= p_i \hat{v}_i(1) + q_i \hat{v}_i(0) \\
\hat{u}_i^{\text{mp,tot}} &= p_i \hat{u}_i^{\text{mp}}(x_1, \dots, x_{i-1}, 1) + q_i \hat{u}_i^{\text{mp}}(x_1, \dots, x_{i-1}, 0).
\end{aligned}$$

The derivative $V'(\alpha)$ can be approximated by the (central) finite difference (FD)

$$V'(\alpha) \approx \frac{V(\alpha + \delta) - V(\alpha - \delta)}{2\delta} = \frac{\tilde{\mathbb{E}}_{\alpha+\delta} [\psi(X)L^2(X)] - \tilde{\mathbb{E}}_{\alpha-\delta} [\psi(X)L^2(X)]}{2\delta} \tag{10}$$

for some small $\delta > 0$, and this finite difference can be estimated by performing two simulations with IS, one at $\alpha + \delta$ and one at $\alpha - \delta$, with common random numbers across the two (i.e., using inversion with the same uniform random numbers to generate the state of each link, for the two simulations [1, 13]). Decreasing δ toward 0 decreases the bias of the derivative estimator (10), but in general it may increase the variance. Then, to ensure convergence of the SA algorithm to α^* , one should use a decreasing sequence $\{\delta_\ell, \ell \geq 0\}$ of values of δ that converges slowly to zero. The appropriate choice of such a sequence and the corresponding convergence rate of SA depend on the behavior of the estimator of $V(\alpha)$ as a function of α , and are studied in [16].

In our case, it turns out that the variance of the FD estimator with common random numbers remains bounded when $\delta \rightarrow 0$. This means that not only we can take δ as small as we want, but we can also take the limit when $\delta \rightarrow 0$ in (10), and we obtain an unbiased derivative estimator that can be computed from a single simulation run. This is an instance of a likelihood ratio (LR) gradient estimator [1, 9]. More specifically, the limit of (10) when $\delta \rightarrow 0$ is the same as the derivative of (9) with respect to α , which gives

$$\begin{aligned}
&\sum_{x \in \{0,1\}^m} \psi(x) \sum_{i=1}^m p_{i,x_i} \frac{\hat{v}_i^{\text{tot}} \hat{u}_i^{\text{mp}}(x_i) - \hat{v}_i(x_i) \hat{u}_i^{\text{mp,tot}}}{(\alpha \hat{v}_i(x_i) + \hat{u}_i^{\text{mp}}(x_i))^2} \prod_{e=1, e \neq i}^m \left(p_{e,x_e} \frac{\alpha \hat{v}_e^{\text{tot}} + \hat{u}_e^{\text{mp,tot}}}{\alpha \hat{v}_e(x_e) + \hat{u}_e^{\text{mp}}(x_e)} \right) \\
&= \tilde{\mathbb{E}}_\alpha \left[\psi(X)L^2(X) \sum_{i=1}^m \frac{\hat{v}_i^{\text{tot}} \hat{u}_i^{\text{mp}}(X_i) - \hat{v}_i(X_i) \hat{u}_i^{\text{mp,tot}}}{(\alpha \hat{v}_i(X_i) + \hat{u}_i^{\text{mp}}(X_i))(\alpha \hat{v}_i^{\text{tot}} + \hat{u}_i^{\text{mp,tot}})} \right].
\end{aligned}$$

This expression can be estimated at any α by simulating from the IS distribution with parameter α , and computing the quantity in brackets as an (unbiased) derivative estimator.

At step ℓ of the SA algorithm, we will simulate n_ℓ independent realizations of this estimator and take the average for $\widehat{V}'(\alpha_\ell)$. How should we choose the n_ℓ 's? Suppose we measure the computing effort (or cost) by the total number of simulation runs. Thus, for a computing budget B , the number of SA iterations that can be performed is $\ell(B) = \max\{\ell : n_1 + \dots + n_\ell \leq B\}$. The variance of the unbiased estimator $\widehat{V}'(\alpha_\ell)$ is proportional to $1/n_\ell$ but its computation time is proportional to n_ℓ . It is shown in [16] that under these circumstances, the variance of the estimator of α^* returned by the SA algorithm (with or without averaging) as a function of the total computing budget B does not depend on the n_ℓ 's, asymptotically when $B \rightarrow \infty$. Roughly, doing n_ℓ SA iterations using a derivative estimator based on a single run is approximately equivalent to doing a single SA iteration using a derivative estimator based on n_ℓ simulation runs [16]. In practice, for a finite computing budget, we recommend taking $n_\ell = 1$. This is what we did in our experiments.

Another (indirect) way of minimizing the variance of the IS estimator with respect to α is minimize the Kullback-Leibler distance between the zero-variance change of measure and the IS parametric family, as explained in [20,21] and references therein. This can also be done via SA, in the same way as for the variance. In our numerical experiments, we found no significant difference between the results from this approach and those based directly on the variance. Therefore, we will report only our results on the latter.

4.4 Summary of the SA Algorithm

The SA algorithm that we have used to estimate the optimal α can be summarized as follows. For the stopping criterion, we specify a minimal number of iterations $\ell_{\min} > \ell_0$, a window size w , and a small preselected *relative* threshold $\epsilon > 0$. At each iteration after the first ℓ_{\min} iterations, we compute the sum of absolute changes on α_ℓ over the last w iterations, and we stop if this sum is less than $\min(\alpha_\ell, 1 - \alpha_\ell) \epsilon$. We use $\min(\alpha_\ell, 1 - \alpha_\ell)$ so that the stopping criterion is based on $1 - \alpha_\ell$ when α_ℓ is close to 1 and on α_ℓ when α_ℓ is close to 0. We also select a maximal number of iterations ℓ_{\max} , after which we stop in all cases.

5 Numerical illustrations

In this section, we report some numerical results obtained with the proposed method on representative examples, and compare them with those based on the mincut-maxprob approximation. For all the numerical experiments reported here, the sample size for the estimation after selecting α is $n = 10^5$. For the heuristic of Section 4.1, we used a sample size of $n_0 = 10^4$ to estimate u in the first stage. In the SA algorithm, we took $w = 1000$, $\epsilon = 10^{-4}$, $\ell_{\min} = 100$, $\ell_{\max} = 10^5$, and $\ell_0 = \ell/2$.

Algorithm 1 The stochastic approximation algorithm to estimate α^*

Select the parameters $\beta, e, C, w, \epsilon, \ell_{\min}, \ell_{\max}$, and ℓ_0 ;
 Select an initial value α_0 ;
 $\ell := 0$;
repeat
 Compute a derivative estimator $\widehat{V}'(\alpha_\ell)$ based on a sample of n_ℓ independent simulation runs with the IS estimator based on (4) with $\alpha = \alpha_\ell$;
 Compute $\alpha_{\ell+1}$ according to (7);
 Compute or update $E = \sum_{j=\ell-w+1}^{\ell} |\alpha_{j+1} - \alpha_j|$;
 $\ell := \ell + 1$;
until ($\ell > \ell_{\min}$ and $E < \min(\alpha_\ell, 1 - \alpha_\ell) \epsilon$) or $\ell \geq \ell_{\max}$;
return $\hat{\alpha} = \hat{\alpha}_{\ell_0, \ell}$.

Example 1 (Continued). We return to the topology described in Figure 1. Table 3 reports the estimate \hat{u} , its empirical RE, and the value $\hat{\alpha}$ of α used for IS, with $\hat{\alpha} = \alpha_{\text{tot}}$ as defined in (6) (heuristic) or $\hat{\alpha}$ obtained from the SA algorithm to minimize the variance (SA).

Table 3 Empirical mean \hat{u} and $\widehat{\text{RE}}$ of the IS estimator of u based on the linear combination method with α selected by the heuristic and with α learned by SA, for the graph of Figure 1, for various values of r , and q chosen so that u is near 10^{-8} .

method	r	q	$10^8 \hat{u}$	$\widehat{\text{RE}}$	$\hat{\alpha}$
heuristic	2	0.00007	1.47	0.22	0.999953
	5	0.020	1.07	0.37	0.9999987
	10	0.1245	1.11	1.6	0.99999969
	30	0.371	1.14	10	0.99999983
	40	0.427	1.21	30	0.999999879
	50	0.4665	1.02	14	0.99999989
	90	0.559	1.28	14	0.999999814
	100	0.575	1.63	34	0.999999873
	200	0.655	1.17	17	0.9999999278
SA	2	0.00007	1.470	0.15	0.9999743
	5	0.020	1.067	0.33	0.999999329
	10	0.1245	1.10	1.3	0.99999987
	30	0.371	1.13	12	0.999999855
	40	0.427	1.15	24	0.9999866
	50	0.4665	1.13	23	0.99999963
	90	0.559	1.43	27	0.99999625
	100	0.575	1.68	30	0.999998
	200	0.655	1.072	11	0.999998587

With the heuristic, α_{tot} is very close to 1 in all cases. This is because $\widehat{u}^{\text{mc}}(\emptyset)$ is very small, $\widehat{u}^{\text{mp}}(\emptyset)$ is close to 1, and u is around 10^{-8} . For large r , $\widehat{u}^{\text{mc}}(\emptyset)$ becomes so small that it has practically no influence on the value of α_{tot} , and we have $1 - \alpha_{\text{tot}} \approx 10^{-8} \times \widehat{u}^{\text{mp}}(\emptyset)$, so the heuristic ends up taking just $\widehat{u}^{\text{mp}}(\emptyset)$ rescaled by a factor of 10^{-8} . Note that to use $\widehat{u}^{\text{mp}}(\emptyset)$ without rescaling, we must take $\alpha = 0$, and not $\alpha = 1$. For smaller r (and small q), e.g., $r \leq 40$, we saw that the mincut-maxprob approximation does better than minpath-maxprob. In this case, both the heuristic and

SA use a more balanced linear combination in which the two terms contribute. In all cases, the heuristic and SA use similar values of α and provide similar results (slightly better for SA for small r). By exploiting the linear combination of two approximations, they provide a more robust approach than either approximation taken alone.

Example 2 We now consider the classical dodecahedron topology, a popular benchmark for this problem, illustrated in Figure 2. It is made of 20 nodes and 30 links, and we want to compute the probability that nodes s and t are disconnected. We consider the homogeneous case, where all links have the same unreliability q . The results are given in Table 4, for four values of q .

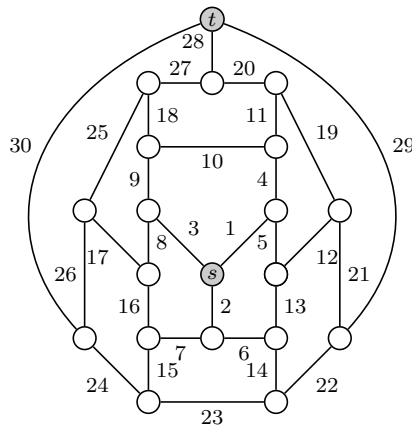


Fig. 2 A dodecahedron graph with 20 nodes and 30 links.

In agreement with [15], the RE with the mincut-maxprob approximation decreases with decreasing q (a vanishing RE occurs when $q \rightarrow 0$). We are in a context where this method is very efficient. As expected, the minpath-maxprob approximation also becomes irrelevant, because some important configurations are not sampled anymore. This is the reason why the estimate \hat{u} is only about half the true value of u with MP for $q = 10^{-3}$ and 10^{-4} . In this particular case, the \widehat{RE} is probably much smaller than the true RE (the empirical results are misleading), and this is due to a very noisy variance estimator. The heuristic gives an acceptable approximation, but it is largely dominated by SA when q is small. Note that when $q \rightarrow 0$, the combined method with SA becomes essentially the same as the mincut-maxprob-based algorithm, as expected. But for larger q , the combined method somehow rescales the approximation to provide a better estimator. Thus, the learning algorithm manages to find the best solution in the considered family in all cases.

Example 3 We now consider the topology described in Figure 3, made of two lines of m nodes between s and t , again with homogeneous links of unreliability q , and such that we want to compute the probability that nodes s and t are disconnected.

Table 4 Estimation of the graph unreliability for the dodecahedron topology described in Figure 2.

method	q	\hat{u}	\widehat{RE}	$\hat{\alpha}$	$\hat{u}^{mc}(\theta)$	$\hat{u}^{mp}(\theta)$
MC	10^{-1}	2.869×10^{-3}	1.91			
	10^{-2}	2.063×10^{-6}	0.175			
	10^{-3}	2.006×10^{-9}	0.050			
	10^{-4}	2.001×10^{-12}	0.016			
MP	10^{-1}	2.907×10^{-3}	6.69			
	10^{-2}	1.667×10^{-6}	85.4			
	10^{-3}	1.003×10^{-9}	0.046			
	10^{-4}	1.000×10^{-12}	0.0136			
heuristic	10^{-1}	2.879×10^{-3}	1.28	0.9955134	10^{-3}	0.4095
	10^{-2}	2.063×10^{-6}	0.348	0.9999783	10^{-6}	0.049
	10^{-3}	2.006×10^{-9}	0.303	0.99999979	10^{-9}	0.00499
	10^{-4}	2.000×10^{-12}	0.298	0.9999999979	10^{-12}	4.999×10^{-4}
SA	10^{-1}	2.870×10^{-3}	0.722	0.99339		
	10^{-2}	2.061×10^{-6}	0.171	0.9999999999999993		
	10^{-3}	2.006×10^{-9}	0.0488	0.9999999999999999		
	10^{-4}	2.001×10^{-12}	0.016	0.9999999999999999		

The results are in Table 5. For this topology, the mincut-maxprob approximation performs poorly, for the same reasons as in Example 1, and the best results are again obtained via SA, although the heuristic also performs well when m is small. One case where the combined method provides a large improvement over mincut-maxprob is when $q = 0.1$ and $m = 20$, where the RE is reduced from 189 to 0.341 with SA; that is, the variance (and the required computing effort for a given precision) is reduced approximately by a factor of $554^2 \approx 307,000$.

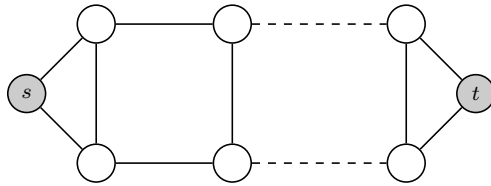


Fig. 3 A topology made of two rows of links and k nodes on each row.

Example 4 Consider an extreme topology made of a single line of $m+1$ nodes and m links, as described in Figure 4, where we want to estimate the probability that the two extreme nodes are connected. Suppose that all the links have the same unreliability q . It is easy to see that the exact unreliability of this system is $u = 1 - (1 - q)^m$.

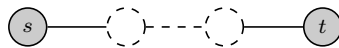


Fig. 4 A topology made of one line of m links.

Table 5 Empirical mean \hat{u} and $\widehat{\text{RE}}$ of the IS estimator of u based on different zero-variance approximations, for the graph of Figure 3, for various values of r , and q chosen so that u is near 10^{-8} .

method	q	m	\hat{u}	$\widehat{\text{RE}}$	$\hat{\alpha}$	$\hat{u}^{\text{mc}}(\theta)$	$\hat{u}^{\text{mp}}(\theta)$
MC	0.5	2	0.672	0.42			
		10	0.995	2.0			
		20	0.998	2.1			
	0.1	2	0.0329	0.50			
		10	0.123	5.3			
		20	0.518	190			
MP	0.5	2	0.672	0.20			
		10	0.9889	0.049			
		20	0.99982	0.0063			
	0.1	2	0.0329	1.5			
		10	0.1208	1.5			
		20	0.2182	1.3			
heuristic	0.5	2	0.672	0.14	0.33	0.25	0.875
		10	0.988	0.044	0.0255	0.25	0.999
		20	0.99978	0.014	0.02614	0.25	0.9999995
	0.1	2	0.03315	0.56	0.912	10^{-2}	0.270
		10	0.1205	0.39	0.836	10^{-2}	0.686
		20	0.2188	0.34	0.8536	10^{-2}	0.8905
SA	0.5	2	0.672	0.13	0.397		
		10	0.9886	0.045	0.0225		
		20	0.99983	0.006	1.6×10^{-36}		
	0.1	2	0.0330	0.33	0.946		
		10	0.1202	0.34	0.877		
		20	0.2184	0.32	0.828		

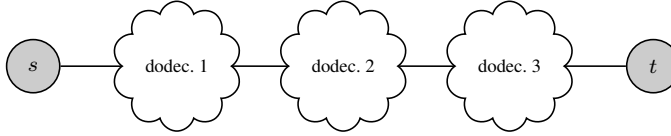
Table 6 shows the results obtained for this topology. A $\widehat{\text{RE}}$ equal to zero indicates that we have a zero-variance estimator. Indeed, here the minpath-maxprob approximation gives u exactly, whereas the mincut-maxprob approximation is not very good because it considers only one mincut out of the m available mincuts (and the numerous other cuts). The learning algorithm with SA adapts itself very well to the characteristics of this topology, and even the heuristic-based method is perfect from (3), because the minpath-maxprob is the exact unreliability.

Example 5 We consider another large example, of a size for which simulation is typically required, but built with a structure such that we are able to know the exact solution. We consider three-dodecahedron topologies in series (so that the reliability of the resulting graph is the cube of the reliability of the dodecahedron), as shown in Figure 5. The source node is the source of the first copy and the destination is the destination of the third one, while the destination of the first (respectively second) copy is the source of the second (respectively third). We still consider the homogeneous case for different unreliability values q . Table 7 gives the results obtained for this topology, for different link unreliabilities.

Here too, the mincut-maxprob approximation improves as $q \rightarrow 0$ while the minpath-maxprob approximation becomes inefficient, to the point of producing here estimates of u and RE that are too small by a large factor when $q \leq 10^{-3}$. The reason is that the important configurations (that contribute to u) are still too rare with this IS scheme and are not sampled. On the other hand, the learning-based algorithm

Table 6 Estimation of u for the topology described in Figure 4.

method	q	m	\hat{u}	\widehat{RE}	$\hat{\alpha}$	$\hat{u}^{mc}(\emptyset)$	$\hat{u}^{mp}(\emptyset)$
MC	10^{-9}	20	2.883×10^{-8}	84			
		40	4.456×10^{-8}	154			
		100	4.456×10^{-8}	154			
MP	10^{-9}	20	1.999×10^{-8}	0			
		40	3.999×10^{-8}	0			
		100	9.999×10^{-8}	0			
heuristic	10^{-9}	20	1.999×10^{-8}	0	0	10^{-9}	1.999×10^{-8}
		40	3.999×10^{-8}	0	0	10^{-9}	3.999×10^{-8}
		100	9.999×10^{-8}	0	0	10^{-9}	9.999×10^{-8}
SA	10^{-9}	20	1.999×10^{-8}	0	0		
		40	3.999×10^{-8}	0	0		
		100	9.999×10^{-8}	0	0		

**Fig. 5** Three dodecahedron topologies in series**Table 7** Estimation of the graph unreliability for the three dodecahedron topologies in series.

method	q	\hat{u}	\widehat{RE}	$\hat{\alpha}$	$\hat{u}^{mc}(\emptyset)$	$\hat{u}^{mp}(\emptyset)$
MC	10^{-1}	8.577×10^{-3}	2.8			
	10^{-2}	6.173×10^{-6}	1.3			
	10^{-3}	6.012×10^{-9}	1.3			
	10^{-4}	5.989×10^{-12}	1.3			
MP	10^{-1}	8.205×10^{-3}	6.8			
	10^{-2}	4.339×10^{-6}	91			
	10^{-3}	1.002×10^{-9}	0.060			
	10^{-4}	1.000×10^{-12}	0.018			
heuristic	10^{-1}	8.584×10^{-3}	0.75	0.990	10^{-3}	0.794
	10^{-2}	6.015×10^{-6}	0.31	0.9999635	10^{-6}	0.140
	10^{-3}	6.015×10^{-9}	0.28	0.999999665	10^{-9}	1.489×10^{-2}
	10^{-4}	5.997×10^{-12}	0.27	0.999999996	10^{-12}	1.498×10^{-3}
SA	10^{-1}	8.599×10^{-3}	0.71	0.991277		
	10^{-2}	6.188×10^{-6}	0.25	0.999974		
	10^{-3}	6.014×10^{-9}	0.22	0.99999975		
	10^{-4}	5.997×10^{-12}	0.22	0.9999999756		

performs well and adapt themselves properly to the graph parameters. Even for small link unreliabilities, it performs better than the mincut-maxprob approximation, contrary to the dodecahedron topology for which we had similar results. The reason for the difference is probably that the number of mincuts with maximal probability is larger on this topology and only one is considered with the mincut-maxprob approximation, hence the approximation is poorer in this case.

6 Conclusions

We have introduced a new approximation of the static unreliability u of a graph, based on minimal paths, called the minpath-maxprob approximation, which underestimates u . This approximation is dual to the mincut-maxprob approximation, which overestimates u . Then we proposed an IS scheme that uses a convex combination of the two approximations to determine the change of measure at each step. This new method has the advantage of adapting itself to various kinds of asymptotic behaviors from which rarity occur, instead of just rarer failure of individual links. In our numerical experiments, this new method was much more robust than taking either of the two non-combined approximations alone.

On the other hand, imposing a single parameter α for all stages, which we did for simplicity, prevents us from mimicking the zero-variance IS as closely as we would like. In our further work, we aim at investigating two directions: (i) instead of taking a single α for all steps i of the algorithm (step i is when we sample link i), we could try to find the optimal vector $(\alpha_1, \dots, \alpha_m)$ such that α_i is used at the i -th step. This would give

$$\widehat{u}_{i+1}(x_1, \dots, x_i) = \alpha_i \widehat{u}_{i+1}^{\text{mc}}(x_1, \dots, x_i) + (1 - \alpha_i) \widehat{u}_{i+1}^{\text{mp}}(x_1, \dots, x_i)$$

instead of (4). This generalization should lead to a better IS measure, but at the expense of a more complicated optimization. (ii) Another option is to replace the *linear combination* in (4) by the *geometric linear combination*

$$\widehat{u}_{i+1}(x_1, \dots, x_i) = (\widehat{u}_{i+1}^{\text{mc}}(x_1, \dots, x_i))^\alpha \times (\widehat{u}_{i+1}^{\text{mp}}(x_1, \dots, x_i))^{1-\alpha},$$

which would better rescale when the two bounds have very different orders of magnitude, like in Example 1. Preliminary experiments with the heuristic of Section 4.1 suggest that this geometric version indeed works better for this example when r is large, but there are other examples where it does not work better.

Acknowledgement

This work has been supported by INRIA's associated team MOCQUASIN, an ARED Région Bretagne project, as well as an NSERC-Canada Discovery Grant and a Canada Research Chair to the third author.

References

1. S. Asmussen and P. W. Glynn. *Stochastic Simulation*. Springer-Verlag, New York, 2007.
2. M. O. Ball. Computational complexity of network reliability analysis: An overview. *IEEE Transactions on Reliability*, 35(3):230–239, 1986.
3. R. Barlow and F. Proschan. *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Wilson, New York, 1975.
4. H. Cancela, M. El Khadiri, and G. Rubino. Rare event analysis by Monte Carlo techniques in static models. In G. Rubino and B. Tuffin, editors, *Rare Event Simulation Using Monte Carlo Methods*, pages 145–170. Wiley, 2009. Chapter 7.
5. I. B. Gertsbakh and Y. Shpunin. *Models of Network Reliability*. CRC Press, Boca Raton, FL, 2010.

6. S. Juneja and P. Shahabuddin. Rare event simulation techniques: An introduction and recent advances. In S. G. Henderson and B. L. Nelson, editors, *Simulation*, Handbooks in Operations Research and Management Science, pages 291–350. Elsevier, Amsterdam, The Netherlands, 2006. Chapter 11.
7. C. Kollman, K. Baggerly, D. Cox, and R. Picard. Adaptive importance sampling on discrete Markov chains. *Annals of Applied Probability*, 9(2):391–412, 1999.
8. H. J. Kushner and G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer Verlag, New York, second edition, 2003.
9. P. L'Ecuyer. A unified view of the IPA, SF, and LR gradient estimation techniques. *Management Science*, 36(11):1364–1383, 1990.
10. P. L'Ecuyer, J. H. Blanchet, B. Tuffin, and P. W. Glynn. Asymptotic robustness of estimators in rare-event simulation. *ACM Transactions on Modeling and Computer Simulation*, 20(1):Article 6, 2010.
11. P. L'Ecuyer, N. Giroux, and P. W. Glynn. Stochastic optimization by simulation: Numerical experiments with the $M/M/1$ queue in steady-state. *Management Science*, 40(10):1245–1261, 1994.
12. P. L'Ecuyer, M. Mandjes, and B. Tuffin. Importance sampling and rare event simulation. In G. Rubino and B. Tuffin, editors, *Rare Event Simulation Using Monte Carlo Methods*, pages 17–38. Wiley, 2009. Chapter 2.
13. P. L'Ecuyer and G. Perron. On the convergence rates of IPA and FDC derivative estimators. *Operations Research*, 42(4):643–656, 1994.
14. P. L'Ecuyer and B. Tuffin. Approximate zero-variance simulation. In *Proceedings of the 2008 Winter Simulation Conference*, pages 170–181. IEEE Press, 2008.
15. P. L'Ecuyer and B. Tuffin. Approximating zero-variance importance sampling in a reliability setting. *Annals of Operations Research*, 189:277–297, 2011.
16. P. L'Ecuyer and G. Yin. Budget-dependent convergence rate for stochastic approximation. *SIAM Journal on Optimization*, 8(1):217–247, 1998.
17. R. Pasupathy and S. Kim. The stochastic root finding problem: Overview, solutions, and open questions. *ACM Transactions on Modeling and Computer Simulation*, 21(3):Article 19, 2011.
18. B. T. Polyak and A. B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, 30:838–855, 1992.
19. G. Rubino and B. Tuffin, editors. *Rare Event Simulation using Monte Carlo Methods*. Wiley, 2009.
20. R. Y. Rubinstein. The cross-entropy method and rare events for maximal cuts and bipartition problems. *ACM Transactions on Modeling and Computer Simulation*, 12(1):27–53, 2002.
21. R. Y. Rubinstein and D. P. Kroese. *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*. Springer-Verlag, 2004.
22. R. Ruppert. Handbook in sequential analysis. In B. K. Ghosh and P. K. Sen, editors, *Stochastic Approximation*, pages 503–529. Marcel Dekker, New York, 1991.
23. J. C. Spall. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. John Wiley, Hoboken, New Jersey, USA, 2003.