Some applications of importance sampling to dependability analysis

Bruno Tuffin
(based on joint works with H. Cancela, M. El Khadiri, P. L’Ecuyer, G. Rubino, S. Saggadi)

INRIA Rennes - Centre Bretagne Atlantique

Paris
2016
Outline

1. Rare events, Static reliability estimation
2. Graph reductions to decrease the work-normalized variance
3. An adaptive ZVIS approximation
4. Combination with Recursive Variance Reduction
   - Recursive Variance Reduction (RVR) algorithm
   - Zero-variance Approximation RVR
5. Conclusions
Outline

1. Rare events, Static reliability estimation
2. Graph reductions to decrease the work-normalized variance
3. An adaptive ZVIS approximation
4. Combination with Recursive Variance Reduction
   - Recursive Variance Reduction (RVR) algorithm
   - Zero-variance Approximation RVR
5. Conclusions
Introduction: rare events and dependability

- in *telecommunication networks*: loss probability of a small unit of information (a packet, or a cell in ATM networks), connectivity of a set of nodes,
- in *dependability analysis*: probability that a system is failed at a given time, availability, mean-time-to-failure,
- in *air control systems*: probability of collision of two aircrafts,
- in *particle transport*: probability of penetration of a nuclear shield,
- in *biology*: probability of some molecular reactions,
- in *insurance*: probability of ruin of a company,
- in *finance*: value at risk (maximal loss with a given probability in a predefined time),
- ...
Robustness properties

- In rare-event simulation models, we often parameterize with a rarity parameter $\epsilon > 0$ such that $\mu = \mathbb{E}[X(\epsilon)] \to 0$ as $\epsilon \to 0$.

- An estimator $X(\epsilon)$ is said to have \textit{bounded relative variance} (or \textit{bounded relative error}) if $\sigma^2(X(\epsilon))/\mu^2(\epsilon)$ is bounded uniformly in $\epsilon$.

- Interpretation: estimating $\mu(\epsilon)$ with a given relative accuracy can be achieved with a bounded number of replications even if $\epsilon \to 0$.

- Weaker property: \textit{asymptotic optimality} (or \textit{logarithmic efficiency}) if $\lim_{\epsilon \to 0} \ln(\mathbb{E}[X^2(\epsilon)])/\ln(\mu(\epsilon)) = 2$.

- Stronger property: \textit{vanishing relative variance}: $\sigma^2(X(\epsilon))/\mu^2(\epsilon) \to 0$ as $\epsilon \to 0$. Asymptotically, we get the zero-variance estimator.

- Other robustness measures exist (based on higher degree moments, on the Normal approximation, on simulation time...).

  L’Ecuyer, Blanchet, T., Glynn, ACM ToMaCS 2010
**Graph model**

- $M$ links can fail independently, *elementary unreliability* $q_e = 1 - r_e$ for edge $e$.
- What is the probability that the set $K$ of (grey) nodes is connected (in the underlying random partial graph of $G$)?
- $X = (X_1, \ldots, X_M)$ (random) *configuration* with $X_e = 1$ if edge $e$ works, 0 otherwise.
- state of the system: $\phi(X)$, where $\phi(X) = 1$ iff $K$ not connected.
- $u = \mathbb{E}[\phi(X)] = \sum_{x \in \{0,1\}^M} \phi(x) \mathbb{P}[X = x]$. 

![Graph model diagram]
Graph model

- $M$ links can fail independently, *elementary unreliability* $q_e = 1 - r_e$ for edge $e$.
- What is the probability that the set $\mathcal{K}$ of (grey) nodes is connected (in the underlying random partial graph of $G$)?
- $X = (X_1, \ldots, X_M)$ (random) *configuration* with $X_e = 1$ if edge $e$ works, 0 otherwise.
- state of the system: $\phi(X)$, where $\phi(X) = 1$ iff $\mathcal{K}$ not connected.
- $u = \mathbb{E}[\phi(X)] = \sum_{x \in \{0,1\}^M} \phi(x) P[X = x]$. 

\[
\begin{align*}
\text{Graph model} \\
\text{M links can fail independently, elementary unreliability } q_e = 1 - r_e \text{ for edge } e. \\
\text{What is the probability that the set } \mathcal{K} \text{ of (grey) nodes is connected (in the underlying random partial graph of } G) ? \\
\text{X} = (X_1, \ldots, X_M) \text{ (random) configuration with } X_e = 1 \text{ if edge } e \text{ works, 0 otherwise.} \\
\text{state of the system: } \phi(X), \text{ where } \phi(X) = 1 \text{ iff } \mathcal{K} \text{ not connected.} \\
u = \mathbb{E}[\phi(X)] = \sum_{x \in \{0,1\}^M} \phi(x) P[X = x].
\end{align*}
\]
Graph model

- $M$ links can fail independently, *elementary unreliability* $q_e = 1 - r_e$ for edge $e$.
- What is the probability that the set $\mathcal{K}$ of (grey) nodes is connected (in the underlying random partial graph of $G$)?
- $X = (X_1, \ldots, X_M)$ (random) *configuration* with $X_e = 1$ if edge $e$ works, 0 otherwise.
- state of the system: $\phi(X)$, where $\phi(X) = 1$ iff $\mathcal{K}$ not connected.
- $u = \mathbb{E}[\phi(X)] = \sum_{x \in \{0,1\}^M} \phi(x) \mathbb{P}[X = x]$.
Graph model

- $M$ links can fail independently, *elementary unreliability* $q_e = 1 - r_e$ for edge $e$.
- What is the probability that the set $\mathcal{K}$ of (grey) nodes is connected (in the underlying random partial graph of $G$)?
- $X = (X_1, \ldots, X_M)$ (random) *configuration* with $X_e = 1$ if edge $e$ works, 0 otherwise.
- state of the system: $\phi(X)$, where $\phi(X) = 1$ iff $\mathcal{K}$ not connected.
- $u = \mathbb{E}[\phi(X)] = \sum_{x \in \{0,1\}^M} \phi(x) \mathbb{P}[X = x]$.

\[
\begin{array}{c}
\text{Bruno Tuffin (INRIA)} \\
\text{IS and dependability analysis} \\
\text{Int. Conf. on Monte Carlo}
\end{array}
\]
Graph model

- $M$ links can fail independently, elementary unreliability $q_e = 1 - r_e$ for edge $e$.
- What is the probability that the set $\mathcal{K}$ of (grey) nodes is connected (in the underlying random partial graph of $G$)?
- $X = (X_1, \ldots, X_M)$ (random) configuration with $X_e = 1$ if edge $e$ works, 0 otherwise.
- state of the system: $\phi(X)$, where $\phi(X) = 1$ iff $\mathcal{K}$ not connected.
- $u = \mathbb{E}[\phi(X)] = \sum_{x \in \{0,1\}^M} \phi(x) P[X = x]$.

\[ \begin{align*}
\end{align*} \]
Graph model

- $M$ links can fail independently, *elementary unreliability* $q_e = 1 - r_e$ for edge $e$.
- What is the probability that the set $\mathcal{K}$ of (grey) nodes is connected (in the underlying random partial graph of $G$)?
- $X = (X_1, \ldots, X_M)$ (random) *configuration* with $X_e = 1$ if edge $e$ works, 0 otherwise.
- state of the system: $\phi(X)$, where $\phi(X) = 1$ iff $\mathcal{K}$ not connected.
- $u = \mathbb{E}[\phi(X)] = \sum_{x \in \{0,1\}^M} \phi(x) P[X = x]$.

![Graph diagram]

Bruno Tuffin (INRIA)
Graph model

- $M$ links can fail independently, *elementary unreliability* $q_e = 1 - r_e$ for edge $e$.
- What is the probability that the set $K$ of (grey) nodes is connected (in the underlying random partial graph of $G$)?
- $X = (X_1, \ldots, X_M)$ (random) *configuration* with $X_e = 1$ if edge $e$ works, $0$ otherwise.
- state of the system: $\phi(X)$, where $\phi(X) = 1$ iff $K$ not connected.
- $u = \mathbb{E}[\phi(X)] = \sum_{x \in \{0,1\}^M} \phi(x) \mathbb{P}[X = x]$.

We have to sum over the $2^M$ configurations.
Crude simulation

- Consider $n$ independent copies $X^{(i)} = (X_1^{(i)}, \ldots, X_m^{(i)})$ of $X$, and compute $Y^{(i)} = \phi(X^{(i)})$.
- The crude estimator of $q$ is then

$$\hat{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y^{(i)}.$$  

- Confidence interval built from the central limit theorem.
- Rarity issue:
  - We assume $q_e \to 0 \forall e$, so that $u \to 0$.
  - The relative error is proportional to

$$\frac{\sqrt{\text{Var}[\hat{Y}_n]}}{\mathbb{E}[Y]} = \frac{\sqrt{u(1-u)}}{u \sqrt{n-1}} \to \infty$$

as $u \to 0$.
  - As a consequence, more and more paths are required to get a specified relative error as $u \to 0$. 

- Idea: sample the links one after the other, with an IS probability that depends on the state of previously sampled links.

- Let \( u_m(x_1, \cdots, x_{m-1}) \), with \( x_i \in \{0, 1\} \), be the unreliability of the graph \( G \) given the states of the links 1 to \( m - 1 \): if \( x_i = 1 \) the link \( i \) is operational, otherwise it is failed.

- Then \( u = u_1() \).

- Sample state of link \( m \), giving 1 with probability:

\[
\tilde{q}_m = u'_m(x_1, \cdots, x_{m-1}) = \frac{q_m u_{m+1}(x_1, \cdots, x_{m-1}, 0)}{(1 - q_m) u_{m+1}(x_1, \cdots, x_{m-1}, 1) + q_m u_{m+1}(x_1, \cdots, x_{m-1}, 0)}.
\]

- Remark (by conditionning) that

\[
u_m(x_1, \cdots, x_{m-1}) = (1 - q_m) u_{m+1}(x_1, \cdots, x_{m-1}, 1) + q_m u_{m+1}(x_1, \cdots, x_{m-1}, 0).
\]

- The resulting unbiased estimator is \( \phi(X) L(X) \), with

\[
L(X) = \prod_{i=1}^{\ell} L_i(x_i) = \prod_{i=1}^{\ell} \left( x_i \frac{1 - q_i}{1 - \tilde{q}_i} + (1 - x_i) \frac{q_i}{\tilde{q}_i} \right).
\]
Where does it come from?

- From the zero-variance IS for a DTMC \((Y_j)_j\) trying to compute

\[
\mu(Y_0) = \sum_{j=1}^{\tau} c(Y_{j-1}, Y_j)
\]

- Use change of probability transitions

\[
\tilde{P}(y, z) = \frac{P(y, z)(c(y, z) + \mu(z))}{\sum_w P(y, w)(c(y, w) + \mu(w))} = \frac{P(y, z)(c(y, z) + \mu(z))}{\mu(y)}
\]

- This yields the \textit{unique} Markov chain implementation of the zero-variance estimator.
Zero-variance estimation and approximation

**Proposition**

**Using this IS, the estimator has zero variance (always yields \( u \)).**

- **Problem:** the \( u_m(\cdot) \) are not known, otherwise no need to simulate.
- **Principle:** approach \( u_m(\cdot) \) by some \( \hat{u}_m(\cdot) \) and use

\[
\tilde{q}_m = \frac{q_m \hat{u}_{m+1}(x_1, \ldots, x_{m-1}, 0)}{q_m \hat{u}_{m+1}(x_1, \ldots, x_{m-1}, 0) + (1 - q_m) \hat{u}_{m+1}(x_1, \ldots, x_{m-1}, 1)}.
\]
Approximation of the zero-variance estimator

- Our proposal: $\hat{u}_m(x_1, \cdots, x_{m-1})$ is the probability of a mincut of the graph with highest probability, given the state of links 1 to $m-1$.
  - A cut (or $K$-cut) is a set of edges such that, if we remove them, the nodes in $K$ are not in the same connected component.
  - A mincut (minimal cut) is a cut that contains no other cut than itself.

- Intuition: the unreliability is the probability of union of all cuts, the most crucial one(s) being the mincut(s) with highest probability.
- Cuts can be obtained in polynomial time.
Results

- In a given state \((x_1, \ldots, x_{m-1})\), we need to determine \(\hat{u}_{m+1}(x_1, \ldots, x_{m-1}, 1)\) and \(\hat{u}_{m+1}(x_1, \ldots, x_{m-1}, 0)\).

- This adds some computational burden, but should substantially reduce the variance.

**Proposition**

*Bounded relative error* proved in general,
*Vanishing relative error* under identified conditions.
Ex: dodecahedron topology, all links with unreliability $\epsilon$

With respect to crude MC, a computational time increase of 16.
Larger networks: 3 dodecahedrons in parallel

- Vanishing relative error observed
- For 3 dodecahedron in series, Bounded relative error observed
- Works very well for such topologies with close to 100 links, and larger.

<table>
<thead>
<tr>
<th>$q_e = \epsilon$</th>
<th>Estimate</th>
<th>95% confidence interval</th>
<th>std dev.</th>
<th>Relative Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-1}$</td>
<td>$2.3573 \times 10^{-8}$</td>
<td>$(2.2496 \times 10^{-8}, 2.4650 \times 10^{-8})$</td>
<td>$5.49 \times 10^{-8}$</td>
<td>2.3</td>
</tr>
<tr>
<td>$5 \times 10^{-2}$</td>
<td>$2.5732 \times 10^{-11}$</td>
<td>$(2.5138 \times 10^{-11}, 2.6327 \times 10^{-11})$</td>
<td>$3.03 \times 10^{-11}$</td>
<td>1.2</td>
</tr>
<tr>
<td>$10^{-2}$</td>
<td>$8.7655 \times 10^{-18}$</td>
<td>$(8.7145 \times 10^{-18}, 8.8165 \times 10^{-18})$</td>
<td>$2.60 \times 10^{-18}$</td>
<td>0.30</td>
</tr>
</tbody>
</table>
Outline

1. Rare events, Static reliability estimation

2. Graph reductions to decrease the work-normalized variance

3. An adaptive ZVIS approximation

4. Combination with Recursive Variance Reduction
   - Recursive Variance Reduction (RVR) algorithm
   - Zero-variance Approximation RVR

5. Conclusions
Improving ZVIS by applying graph reductions when sampling links

- Each time a link state is generated by the ZVIS algorithm, the graph evolves according to these rules: at step $i$ ($1 \leq i \leq \ell$),
  - either $X_i = 0$ which means that the link is removed,
  - or $X_1 = 1$ which means that the link is fixed, and can then be removed by merging the two nodes it links.

- At each step, we can therefore search if graph reductions can be applied, in order to simplify the topology, and potentially gain in terms of
  - variance
  - computational time (because the size of the graph is smaller).
Considered graph reductions

- **Series reduction:**
  - If node $s \in \mathcal{N}$ has only two incident links, $l_1$ and $l_2$, connecting it to nodes $s_1$ and $s_2$ respectively.
  - If $s \notin \mathcal{K}$, node $s$ can be removed and links $l_1$ and $l_2$ merged into a single one, with unreliability $q = 1 - (1 - q_{l_1})(1 - q_{l_2})$.

  

  \[
  q_1 q_2 \Rightarrow 1 - (1 - q_1)(1 - q_2)
  \]

  - The case $s \in \mathcal{K}$ can hardly be treated without further topology information.

- **Parallel reduction:**
  - If there are two (parallel) links $l_1$ and $l_2$ both connecting nodes $s_1$ and $s_2$.
  - Those two links merged into a single one, with unreliability $q = q_{l_1} q_{l_2}$.

  

  \[
  q_1 q_2 \Rightarrow q_1 q_2
  \]
Two possible combinations with ZVIS

- **Posterior reduction (PR)**
  - link $i$ sampled with failed probability

\[
\hat{q}_i^{(1)} = \frac{q_i \hat{u}_{i+1}(G'_i, 0)}{q_i \hat{u}_{i+1}(G'_i, 0) + (1 - q_i) \hat{u}_{i+1}(G'_i, 1)},
\]

where $G'_i$ graph resulting from previous link samplings and reductions

- link $i$ is removed if $X_i = 0$ and compressed if $X_i = 1$
- new reductions are searched, leading to a new graph $G'_{i+1}$.

- **Look-ahead reduction (LAR)**
  - the probability that $i$ is failed:

\[
\hat{q}_i^{(2)} = \frac{q_i \hat{u}_{i+1}(G'_{i}, 0)}{q_i \hat{u}_{i+1}(G'_{i}, 0) + (1 - q_i) \hat{u}_{i+1}(G'_{i}, 1)},
\]

where $G'_{i,k}$ for $k \in \{0, 1\}$ is the graph reduced after setting $X_i = k$

- This requires to make two copies of the graph, setting $X_i = 0$ for the first and $X_i = 1$ for the other,
- those two resulting graphs being reduced according to the above rules
- When link $i$ effectively sampled, we choose the appropriate already reduced graph.
Expected gain

- **Computational time:**
  - Time for graph reduction searches and making copies of the graph
  - but it decreases the number of links to sample and the number of mincut-maxprob approximations to be computed.

- **Variance:**
  - better mincut-maxprob approximation of the graph unreliabilities at the different steps, *usually* resulting in smaller variance.

- **Comparing the two implementations:**
  - LAR requires additional time to make copies of the graph and to perform twice more reductions at any given step
  - but computing the mincut-maxprob on an already reduced graph takes a shorter time than before proceeding to a reduction.
  - Moreover we usually get a better approximation of the zero-variance IS with this procedure.
First sample link 1.
- If $X_1 = 1$, 

Toy example with cascading reductions
First sample link 1.

- If $X_1 = 1$,
  - the graph can then be reduced by compressing link 1, merging nodes $A$ and $B$, 

Toy example with cascading reductions
First sample link 1.
- If $X_1 = 1$,
  - the graph can then be reduced by compressing link 1, merging nodes $A$ and $B$,
  - then a parallel reduction of links 2 and 3 can be applied.
First sample link 1.

- If $X_1 = 1$,
  - the graph can then be reduced by compressing link 1, merging nodes $A$ and $B$,
  - then a parallel reduction of links 2 and 3 can be applied.
  - This new link is then in series with link 5, leading to a reduction.
First sample link 1.

- If $X_1 = 1$,
  - the graph can then be reduced by compressing link 1, merging nodes $A$ and $B$,
  - then a parallel reduction of links 2 and 3 can be applied.
  - This new link is then in series with link 5, leading to a reduction.
  - The resulting graph is then just made of two parallel links which can therefore be reduced.
First sample link 1.
- If $X_1 = 1$,
  - the graph can then be reduced by compressing link 1, merging nodes $A$ and $B$,
  - then a parallel reduction of links 2 and 3 can be applied.
  - This new link is then in series with link 5, leading to a reduction.
  - The resulting graph is then just made of two parallel links which can therefore be reduced.
  - By IS, the link is necessarily considered failed. Just one link sampled!
If $X_1 = 0$, proceeding similarly,
- Link 1 is removed.
- Links 3 and 4 are then in series and can be reduced,
- the resulting link becomes a parallel link with link 5, reduced
- to a link in series with link 2, which can be reduced to lead to a single link, necessarily failed under IS. Just one link sampled too!
In terms of variance

The algorithms have the following robustness properties, as failures of individual links go to zero:

- On our toy example:
  - With PR, VRE is obtained
  - While with LAR, zero variance is obtained (perfect approximation of unreliabilities).

- With full generality,

Proposition

*Our algorithms satisfy BRE.*
\[ q_i = \epsilon \; \forall i \]

<table>
<thead>
<tr>
<th>Met.</th>
<th>$\epsilon$</th>
<th>Variance</th>
<th>RE.</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZVIS</td>
<td>$10^{-1}$</td>
<td>$1.1048 \times 10^{-5}$</td>
<td>1.1733</td>
<td>15.18</td>
</tr>
<tr>
<td></td>
<td>$10^{-2}$</td>
<td>$1.1670 \times 10^{-13}$</td>
<td>0.1652</td>
<td>14.35</td>
</tr>
<tr>
<td></td>
<td>$10^{-3}$</td>
<td>$1.2714 \times 10^{-20}$</td>
<td>0.0561</td>
<td>14.88</td>
</tr>
<tr>
<td>PR</td>
<td>$10^{-1}$</td>
<td>$5.5452 \times 10^{-6}$</td>
<td>0.8190</td>
<td>12.14</td>
</tr>
<tr>
<td></td>
<td>$10^{-2}$</td>
<td>$9.8889 \times 10^{-14}$</td>
<td>0.1522</td>
<td>15.33</td>
</tr>
<tr>
<td></td>
<td>$10^{-3}$</td>
<td>$9.5548 \times 10^{-21}$</td>
<td>0.0487</td>
<td>13.87</td>
</tr>
<tr>
<td>LAR</td>
<td>$10^{-1}$</td>
<td>$3.9203 \times 10^{-6}$</td>
<td>0.6880</td>
<td>10.29</td>
</tr>
<tr>
<td></td>
<td>$10^{-2}$</td>
<td>$4.4955 \times 10^{-14}$</td>
<td>0.1028</td>
<td>7.48</td>
</tr>
<tr>
<td></td>
<td>$10^{-3}$</td>
<td>$2.4094 \times 10^{-21}$</td>
<td>0.0244</td>
<td>7.55</td>
</tr>
</tbody>
</table>
Outline

1. Rare events, Static reliability estimation
2. Graph reductions to decrease the work-normalized variance
3. An adaptive ZVIS approximation
4. Combination with Recursive Variance Reduction
   - Recursive Variance Reduction (RVR) algorithm
   - Zero-variance Approximation RVR
5. Conclusions
Limits of the above ZVIS approximation

- Shown to be very efficient for very low link unreliabilities
- But system failure rarity may come from other reasons. Ex: large number of possible paths.

Increasing $k$ but keeping the same overall unreliability

<table>
<thead>
<tr>
<th>$k$</th>
<th>$q_e$</th>
<th>$10^8 \hat{u}$</th>
<th>$\hat{R}E$</th>
<th>$\hat{u}_1^{mc}(\emptyset) = q^r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$7 \times 10^{-5}$</td>
<td>1.46</td>
<td>0.33</td>
<td>$4.9 \times 10^{-9}$</td>
</tr>
<tr>
<td>5</td>
<td>0.02</td>
<td>1.06</td>
<td>0.46</td>
<td>$3.2 \times 10^{-9}$</td>
</tr>
<tr>
<td>10</td>
<td>0.1245</td>
<td>1.11</td>
<td>1.8</td>
<td>$8.9 \times 10^{-10}$</td>
</tr>
<tr>
<td>30</td>
<td>0.371</td>
<td>1.14</td>
<td>7.9</td>
<td>$1.2 \times 10^{-13}$</td>
</tr>
<tr>
<td>40</td>
<td>0.427</td>
<td>1.05</td>
<td>9.9</td>
<td>$1.6 \times 10^{-15}$</td>
</tr>
<tr>
<td>50</td>
<td>0.4665</td>
<td>1.08</td>
<td>31</td>
<td>$2.7 \times 10^{-17}$</td>
</tr>
<tr>
<td>70</td>
<td>0.521</td>
<td>1.35</td>
<td>22</td>
<td>$1.5 \times 10^{-20}$</td>
</tr>
<tr>
<td>100</td>
<td>0.575</td>
<td>1.48</td>
<td>40</td>
<td>$9.2 \times 10^{-25}$</td>
</tr>
<tr>
<td>200</td>
<td>0.655</td>
<td>0.48</td>
<td>44</td>
<td>$1.8 \times 10^{-37}$</td>
</tr>
</tbody>
</table>
Minpath-based approximation

- **Path**: set $P$ of links such that when up, the nodes in $\mathcal{K}$ are connected.
- **Minpath**: path with no strict subset that is a path.
- **Minpath-maxprob approximation**: max probability of a minpath, $\hat{u}_{\text{mp}}(G) = 1 - \max_{P \in \mathcal{F}_G} p(P)$.
- Computed thanks to Dijkstra algorithm.
- Replacing the mincut-maxprob approximation in ZVIS

\[ \begin{array}{|c|c|c|c|c|}
\hline
k & q_e & 10^8\hat{u} & \hat{R} & \hat{u}_{\text{mp}}(\emptyset) \\
\hline
2 & 0.00007 & 1.68 & 66 & 0.0002 \\
5 & 0.02 & 3.18 & 160 & 0.058 \\
10 & 0.1245 & 1.15 & 110 & 0.32 \\
30 & 0.371 & 1.36 & 75 & 0.75 \\
40 & 0.427 & 1.20 & 36 & 0.81 \\
50 & 0.4665 & 0.98 & 26 & 0.84 \\
70 & 0.521 & 1.58 & 17 & 0.89 \\
90 & 0.559 & 1.19 & 6.6 & 0.91 \\
100 & 0.575 & 1.52 & 9.8 & 0.92 \\
200 & 0.655 & 1.13 & 3.9 & 0.95 \\
\hline
\end{array} \]
What if we combine both approximations?

- Indeed, $\hat{u}^{mc} \leq u \leq \hat{u}^{mp}$.
- Take at each step

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

- Should always be closer to the unreliability.
- How to determine the best $\alpha$?
- First heuristic:
  - Compute a rough estimate $\hat{u}_{n_0}(G)$ of $u$
  - Take

$$\alpha = \alpha_{tot} \overset{\text{def}}{=} \frac{\hat{u}^{mp}(\emptyset) - \hat{u}_{n_0}(G)}{\hat{u}^{mp}(\emptyset) - \hat{u}^{mc}(\emptyset)},$$

the $\alpha$ leading to the above equality with this rough estimate. for the full network unreliability.
Learning through a Robbins-Monro algorithm

- Goal: compute the $\alpha$ minimizing the variance, i.e., st $V'(\alpha) = 0$.

1. $\ell = 0$, start with a $\alpha_0$ (the one from the heuristic)
   1. $\ell = \ell + 1$
   2. estimate $\hat{V}'(\alpha_\ell)$
   3. Stop when it seems to have converged, or update again.

2. Launch the real simulation with the last $\alpha_\ell$.

- I skip the computation of the derivative and choice of parameters (paper available on requests).
Ex: transport network of ANTEL
<table>
<thead>
<tr>
<th>method</th>
<th>$q$</th>
<th>$\hat{\mu}$</th>
<th>$\hat{R}_E$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\mu}^{\text{mc}}(\emptyset)$</th>
<th>$\hat{\mu}^{\text{mp}}(\emptyset)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC</td>
<td>$10^{-2}$</td>
<td>$1.22 \times 10^{-2}$</td>
<td>108</td>
<td>108</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$10^{-3}$</td>
<td>$2.11 \times 10^{-4}$</td>
<td>103</td>
<td>103</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$10^{-4}$</td>
<td>$1.92 \times 10^{-6}$</td>
<td></td>
<td>109</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MP</td>
<td>$10^{-2}$</td>
<td>$7.58 \times 10^{-3}$</td>
<td>2.5</td>
<td>2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$10^{-3}$</td>
<td>$7.48 \times 10^{-5}$</td>
<td>8.4</td>
<td>8.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$10^{-4}$</td>
<td>$6.74 \times 10^{-7}$</td>
<td>25</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>heuristic</td>
<td>$10^{-2}$</td>
<td>$7.49 \times 10^{-3}$</td>
<td>2.7</td>
<td>2.7</td>
<td>0.92873</td>
<td>$10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>$10^{-3}$</td>
<td>$7.37 \times 10^{-5}$</td>
<td>11</td>
<td>11</td>
<td>0.977225</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>$10^{-4}$</td>
<td>$7.25 \times 10^{-7}$</td>
<td>7.2</td>
<td>7.2</td>
<td>0.99770</td>
<td>$10^{-8}$</td>
</tr>
<tr>
<td>SA</td>
<td>$10^{-2}$</td>
<td>$7.54 \times 10^{-3}$</td>
<td>1.9</td>
<td>1.9</td>
<td>0.59887</td>
<td>$10^{-8}$</td>
</tr>
<tr>
<td></td>
<td>$10^{-3}$</td>
<td>$7.27 \times 10^{-5}$</td>
<td>2.8</td>
<td>2.8</td>
<td>0.99838</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>$10^{-4}$</td>
<td>$7.26 \times 10^{-7}$</td>
<td>2.8</td>
<td>2.8</td>
<td>0.999843</td>
<td>$1.998 \times 10^{-3}$</td>
</tr>
</tbody>
</table>
Outline

1. Rare events, Static reliability estimation
2. Graph reductions to decrease the work-normalized variance
3. An adaptive ZVIS approximation
4. Combination with Recursive Variance Reduction
   - Recursive Variance Reduction (RVR) algorithm
   - Zero-variance Approximation RVR
5. Conclusions
Recursive Variance Reduction (RVR)

- Principle: select a $\mathcal{K}$-cutset, i.e., a set $C$ of links whose failure ensures the system failure.

If all links in $C$ are failed (probability $q_C$), the system is failed. Clearly, $q_C \leq q$.

$B_j = \text{“the } j - 1 \text{ first links of } C \text{ are down, but the } j\text{-th is up”}$

$\mathbb{P}[B_j] = (\prod_{k=1}^{j-1} q_k) r_j$

Define $p_j = \mathbb{P}[B_j | \text{at least one link is working}] = \mathbb{P}[B_j]/(1 - q_C)$
Recursive Variance Reduction (RVR)

The RVR estimator:

- Select a cut, and compute $q_C$ and the $p_j$s.
- Pick an edge at random in $C$ according to the probability distribution $(p_j)_{j=1,\ldots,|C|}$.
- Let the chosen edge be the $j$th. Call $G_j$ the graph obtained from $G$ by deleting the first $j - 1$ edges of $C$ and by contracting the $j$th.
- The value $y_{RVR}$ returned by the RVR estimator of $q(G)$, the unreliability of $G$, is recursively defined as

$$y_{RVR}(G) = q_C + (1 - q_C)y_{RVR}(G_j).$$
RVR estimator

Formally, the RVR estimator of \( q(\mathcal{G}) \) is the random variable

\[
Y_{RVR} = q_C + (1 - q_C) \sum_{j=1}^{\vert C \vert} \frac{1_{B_j}}{1 - q_C} Y_{RVR}(\mathcal{G}_j).
\]

Theorem

The estimator is unbiased: \( \mathbb{E}[Y_{RVR}] = q(\mathcal{G}) = q \).

Second moment computed as

\[
\mathbb{E}[Y_{RVR}^2] = q_C^2 + 2q_C(1 - q_C) \left( \sum_{j=1}^{\vert C \vert} \frac{\mathbb{P}[B_j]}{1 - q_C} \mathbb{E}[Y_{RVR}(\mathcal{G}_j)] \right) + (1 - q_C)^2 \left( \sum_{j=1}^{\vert C \vert} \frac{\mathbb{P}[B_j]}{1 - q_C} \mathbb{E}[Y_{RVR}^2(\mathcal{G}_j)] \right).
\]

But no BRE as \( \epsilon \to 0 \).
Zero-variance Approximation RVR

- **Zero-variance change of measure**: chooses the appropriate (ideally the best) IS for the first working link on the cut:
- Choose $B'_j$ with probability $\tilde{p}_j$ in the IS estimator, with

$$
\tilde{p}_j = \frac{P[B_j]q(G_j)}{\sum_{j=1}^{\lvert C \rvert} P[B_k]q(G_k)}
$$

(1)

- Resulting estimator:

$$
Y_{ZRVR} = q_C + \left( \sum_{k=1}^{\lvert C \rvert} P[B_k]q(G_k) \right) \sum_{j=1}^{\lvert C \rvert} 1_{B'_j(G)} \frac{1}{q(G_j)} Y_{ZRVR}(G_j).
$$

**Theorem**

$Y_{ZRVR}$ has variance $\text{Var}[Y_{ZRVR}] = 0$.

- Implementing it requires the knowledge of the $q(G_i)$, but in that case, no need to simulate!
Zero Variance Approximation

- Instead, use some approximation \( \hat{q}(G_i) \) of \( q(G_i) \) plugged into (1).

\[
Y_{AZRVR} = q_c + \left( \sum_{k=1}^{\lvert C \rvert} \mathbb{P}[B_k] \hat{q}(G_k) \right) \sum_{j=1}^{\lvert C \rvert} \mathbf{1}_{B'_j(G)} \frac{1}{\hat{q}(G_j)} Y_{AZRVR}(G_j).
\]

Proposition

If \( \forall 1 \leq j \leq \lvert C \rvert, \hat{q}(G_j) = \Theta(q(G_j)) \) as \( \epsilon \to 0 \), \( Y_{AZRVR} \) verifies BRE property.

- Define the mincut-maxprob approximation \( \hat{q}(G) \) of \( q(G) \) as maximal probability of a mincut of graph \( G \) (computed in polynomial time).

Proposition

With the mincut-maxprob approximation, \( \hat{q}(G_j) = \Theta(q(G_j)) \) as \( \epsilon \to 0 \), therefore BRE property is obtained.

Proposition

If, \( \hat{q}(G_j) = q(G_j) + o(q(G_j)) \) as \( \epsilon \to 0 \) for all \( 1 \leq j \leq \lvert C \rvert \), the Vanishing relative (VRE) property (the RE tends to 0, stronger than just being bounded) is verified.
Three topologies: arpanet, C6, dodecahedron

Bruno Tuffin (INRIA)
<table>
<thead>
<tr>
<th>Network ( (q_e) )</th>
<th>( Q(G) )</th>
<th>( N \times \text{Var}(SMC) )</th>
<th>( N \times \text{Var}(RVR) )</th>
<th>( N \times \text{Var}(AZV) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arp ( (5.00 \times 10^{-1}) )</td>
<td>9.63989 ( \times 10^{-1} )</td>
<td>3.47133 ( \times 10^{-2} )</td>
<td>3.71795 ( \times 10^{-3} )</td>
<td>1.69321 ( \times 10^{-1} )</td>
</tr>
<tr>
<td>Arp ( (3.00 \times 10^{-1}) )</td>
<td>6.81507 ( \times 10^{-1} )</td>
<td>2.17055 ( \times 10^{-1} )</td>
<td>4.74801 ( \times 10^{-2} )</td>
<td>8.45549 ( \times 10^{-1} )</td>
</tr>
<tr>
<td>Arp ( (1.00 \times 10^{-1}) )</td>
<td>9.54229 ( \times 10^{-2} )</td>
<td>8.63174 ( \times 10^{-3} )</td>
<td>1.46865 ( \times 10^{-2} )</td>
<td>9.55806 ( \times 10^{-2} )</td>
</tr>
<tr>
<td>Arp ( (1.00 \times 10^{-2}) )</td>
<td>6.54074 ( \times 10^{-4} )</td>
<td>6.53646 ( \times 10^{-6} )</td>
<td>1.63753 ( \times 10^{-5} )</td>
<td>3.06912 ( \times 10^{-6} )</td>
</tr>
<tr>
<td>Arp ( (1.00 \times 10^{-3}) )</td>
<td>6.05581 ( \times 10^{-6} )</td>
<td>6.05577 ( \times 10^{-6} )</td>
<td>1.60407 ( \times 10^{-5} )</td>
<td>3.43246 ( \times 10^{-6} )</td>
</tr>
<tr>
<td>Arp ( (1.00 \times 10^{-4}) )</td>
<td>6.00560 ( \times 10^{-8} )</td>
<td>6.00560 ( \times 10^{-8} )</td>
<td>1.60041 ( \times 10^{-7} )</td>
<td>3.47090 ( \times 10^{-8} )</td>
</tr>
<tr>
<td>Arp ( (1.00 \times 10^{-5}) )</td>
<td>6.00056 ( \times 10^{-10} )</td>
<td>6.00056 ( \times 10^{-10} )</td>
<td>1.60004 ( \times 10^{-9} )</td>
<td>3.47477 ( \times 10^{-10} )</td>
</tr>
<tr>
<td>Arp ( (1.00 \times 10^{-6}) )</td>
<td>6.00006 ( \times 10^{-12} )</td>
<td>6.00006 ( \times 10^{-12} )</td>
<td>1.60000 ( \times 10^{-11} )</td>
<td>3.47512 ( \times 10^{-12} )</td>
</tr>
<tr>
<td>C6 ( (5.00 \times 10^{-1}) )</td>
<td>7.64160 ( \times 10^{-2} )</td>
<td>7.05766 ( \times 10^{-2} )</td>
<td>7.72612 ( \times 10^{-3} )</td>
<td>7.27858 ( \times 10^{-3} )</td>
</tr>
<tr>
<td>C6 ( (3.00 \times 10^{-1}) )</td>
<td>5.26728 ( \times 10^{-3} )</td>
<td>5.23953 ( \times 10^{-3} )</td>
<td>2.56429 ( \times 10^{-4} )</td>
<td>2.27577 ( \times 10^{-4} )</td>
</tr>
<tr>
<td>C6 ( (1.00 \times 10^{-1}) )</td>
<td>2.00766 ( \times 10^{-5} )</td>
<td>2.00762 ( \times 10^{-5} )</td>
<td>1.28070 ( \times 10^{-6} )</td>
<td>1.17223 ( \times 10^{-6} )</td>
</tr>
<tr>
<td>C6 ( (1.00 \times 10^{-2}) )</td>
<td>2.00001 ( \times 10^{-10} )</td>
<td>2.00001 ( \times 10^{-10} )</td>
<td>1.01244 ( \times 10^{-12} )</td>
<td>1.00225 ( \times 10^{-12} )</td>
</tr>
<tr>
<td>C6 ( (1.00 \times 10^{-3}) )</td>
<td>2.00000 ( \times 10^{-15} )</td>
<td>2.00000 ( \times 10^{-15} )</td>
<td>1.00102 ( \times 10^{-18} )</td>
<td>1.00002 ( \times 10^{-18} )</td>
</tr>
<tr>
<td>C6 ( (1.00 \times 10^{-4}) )</td>
<td>2.00000 ( \times 10^{-20} )</td>
<td>2.00000 ( \times 10^{-20} )</td>
<td>1.00000 ( \times 10^{-23} )</td>
<td>1.00000 ( \times 10^{-23} )</td>
</tr>
<tr>
<td>C6 ( (1.00 \times 10^{-5}) )</td>
<td>2.00000 ( \times 10^{-25} )</td>
<td>2.00000 ( \times 10^{-25} )</td>
<td>1.42434 ( \times 10^{-28} )</td>
<td>1.42434 ( \times 10^{-28} )</td>
</tr>
<tr>
<td>Dod ( (5.00 \times 10^{-1}) )</td>
<td>7.09745 ( \times 10^{-1} )</td>
<td>2.06007 ( \times 10^{-1} )</td>
<td>1.57246 ( \times 10^{-2} )</td>
<td>1.34634 ( \times 10^{-1} )</td>
</tr>
<tr>
<td>Dod ( (3.00 \times 10^{-1}) )</td>
<td>1.68518 ( \times 10^{-1} )</td>
<td>1.40120 ( \times 10^{-1} )</td>
<td>9.22721 ( \times 10^{-3} )</td>
<td>1.68222 ( \times 10^{-2} )</td>
</tr>
<tr>
<td>Dod ( (1.00 \times 10^{-1}) )</td>
<td>2.87960 ( \times 10^{-3} )</td>
<td>2.87131 ( \times 10^{-3} )</td>
<td>5.80985 ( \times 10^{-5} )</td>
<td>6.32871 ( \times 10^{-5} )</td>
</tr>
<tr>
<td>Dod ( (1.00 \times 10^{-2}) )</td>
<td>2.06189 ( \times 10^{-6} )</td>
<td>2.06189 ( \times 10^{-6} )</td>
<td>2.17456 ( \times 10^{-8} )</td>
<td>1.12133 ( \times 10^{-8} )</td>
</tr>
<tr>
<td>Dod ( (1.00 \times 10^{-3}) )</td>
<td>2.00602 ( \times 10^{-9} )</td>
<td>2.00602 ( \times 10^{-9} )</td>
<td>2.01614 ( \times 10^{-11} )</td>
<td>1.01110 ( \times 10^{-11} )</td>
</tr>
<tr>
<td>Dod ( (1.00 \times 10^{-4}) )</td>
<td>2.00060 ( \times 10^{-12} )</td>
<td>2.00060 ( \times 10^{-12} )</td>
<td>2.00160 ( \times 10^{-14} )</td>
<td>1.00110 ( \times 10^{-14} )</td>
</tr>
<tr>
<td>Dod ( (1.00 \times 10^{-5}) )</td>
<td>2.00006 ( \times 10^{-15} )</td>
<td>2.00006 ( \times 10^{-15} )</td>
<td>2.00016 ( \times 10^{-17} )</td>
<td>1.00011 ( \times 10^{-17} )</td>
</tr>
<tr>
<td>Dod ( (1.00 \times 10^{-6}) )</td>
<td>2.00001 ( \times 10^{-18} )</td>
<td>2.00001 ( \times 10^{-18} )</td>
<td>2.00002 ( \times 10^{-20} )</td>
<td>1.00001 ( \times 10^{-20} )</td>
</tr>
</tbody>
</table>
Outline

1. Rare events, Static reliability estimation
2. Graph reductions to decrease the work-normalized variance
3. An adaptive ZVIS approximation
4. Combination with Recursive Variance Reduction
   - Recursive Variance Reduction (RVR) algorithm
   - Zero-variance Approximation RVR
5. Conclusions
A summary of best existing methods for static reliability estimation on the dodecahedron

Without presenting all implementations.

(Normalized) relative error $\frac{\sqrt{n} \times RE}{c_\alpha}$ for various methods and unreliabilities $\epsilon$ of links on the dodecahedron topology

<table>
<thead>
<tr>
<th>Method</th>
<th>$\epsilon = 0.1$</th>
<th>$\epsilon = 10^{-2}$</th>
<th>$\epsilon = 10^{-3}$</th>
<th>$\epsilon = 10^{-4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditioning, Fishman 86</td>
<td>2.6 e+00</td>
<td>1.3 e+00</td>
<td>4.3 e−01</td>
<td>1.4 e−02</td>
</tr>
<tr>
<td>GS Botev et al. 13</td>
<td>4.0 e+00</td>
<td>6.2 e+00</td>
<td>7.7 e+00</td>
<td>8.9 e+00</td>
</tr>
<tr>
<td>Splitting, Murray et al. 13</td>
<td>4.6 e+00</td>
<td>7.1 e+00</td>
<td>8.6 e+00</td>
<td>8.8 e+00</td>
</tr>
<tr>
<td>Permutation MC Gerbatsh</td>
<td>3.0 e+00</td>
<td>4.2 e+00</td>
<td>4.3 e+00</td>
<td>4.4 e+00</td>
</tr>
<tr>
<td>IS: ZVA 2010</td>
<td>1.2 e+00</td>
<td>1.7 e−01</td>
<td>5.7 e−02</td>
<td>1.7 e−02</td>
</tr>
<tr>
<td>RVR Cancela, Khadiri 1995</td>
<td>8.4 e−01</td>
<td>7.1 e−01</td>
<td>7.1 e−01</td>
<td>7.1 e−01</td>
</tr>
<tr>
<td>IS+ RVR: BRD 14</td>
<td>9.5 e−01</td>
<td>7.0 e−01</td>
<td>7.1 e−01</td>
<td>7.1 e−01</td>
</tr>
<tr>
<td>IS+RVR: AZVRD 14</td>
<td>2.8 e−01</td>
<td>5.1 e−02</td>
<td>1.6 e−02</td>
<td>5.0 e−03</td>
</tr>
</tbody>
</table>
Work in progress

- Railway Data Communication System (DCS), with failing nodes

- Dependability including logistics: return to a dynamic model. Two challenges
  - Non-Markovian model
  - more complicated assumptions with logistics on repair teams, spares.
Example: Highly Reliable Markovian Systems (HRMS)

- System with \( c \) types of components. \( Y = (Y_1, \ldots, Y_c) \) with \( Y_i \) number of up components.

- \( \mathbf{1} \): state with all components up.

- Markov chain. Failure rates are \( O(\varepsilon) \), but not repair rates. Failure propagations possible.

- System down when in grey state(s) (in \( \Delta \)).

- Goal: compute \( \mu(y) \) probability to hit \( \Delta \) before \( \mathbf{1} \).

- \( \mu(\mathbf{1}) \) important in dependability analysis,

- Small if \( \varepsilon \) small.
Failure rates are $O(\varepsilon)$, but not repair rates. Failure propagations possible.

Simulation using the embedded DTMC. Failure probabilities are $O(\varepsilon)$ (except from 1). How to improve (accelerate) this?

Existing method: $\forall y \neq 1$, increase the probability of the set of failures to constant $0.5 < q < 0.9$ and use individual probabilities proportional to the original ones (SFB), or uniformly (BFB).

Failures not rare anymore. BRE property verified for BFB.
HRMS Example, and IS

Figure: Original probabilities

Figure: Probabilities under IS/BFB
Complicates the previous model due to the multidimensional description of a state.

The idea is to approach $\mu(y)$ by the probability of the path from $y$ to $\Delta$ with the largest probability.

Intuition: as $\epsilon \to 0$, we get a good idea of the probability.

**Proposition**

*Bounded Relative Error proved (as $\epsilon \to 0$) in general.*

*Even Vanishing Relative Error if $\hat{\mu}(y)$ contains all the paths with the smallest degree in $\epsilon$.*

Other simple version: approach $\mu(y)$ by the (sum of) probability of paths from $y$ with only failure components of a given type.

Gain of several orders of magnitudes + stability of the results with respect to the literature.
HRMS: numerical illustrations

- Comparison of BFB and Zero-Variance Approximation (ZVA).
- $c = 3$ types of components, $n_i$ of type $i$
- $\lambda_1 = \varepsilon$, $\lambda_2 = 1.5\varepsilon$, and $\lambda_3 = 2\varepsilon^2$, $\mu = 1$
- System is down whenever fewer than two components of any one type are operational.

| $n_i$ | $\varepsilon$ | $\mu_0$ | BFB est | ZVA est | BFB $\sigma^2$ | ZVA $\sigma^2$
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.001</td>
<td>$2.6 \times 10^{-3}$</td>
<td>$2.7 \times 10^{-3}$</td>
<td>$2.6 \times 10^{-3}$</td>
<td>$6.2 \times 10^{-5}$</td>
<td>$2.2 \times 10^{-8}$</td>
</tr>
<tr>
<td>6</td>
<td>0.01</td>
<td>$1.8 \times 10^{-7}$</td>
<td>$1.9 \times 10^{-7}$</td>
<td>$1.8 \times 10^{-7}$</td>
<td>$6.3 \times 10^{-11}$</td>
<td>$2.0 \times 10^{-14}$</td>
</tr>
<tr>
<td>6</td>
<td>0.001</td>
<td>$1.7 \times 10^{-11}$</td>
<td>$1.8 \times 10^{-11}$</td>
<td>$1.7 \times 10^{-11}$</td>
<td>$8.8 \times 10^{-19}$</td>
<td>$1.2 \times 10^{-23}$</td>
</tr>
<tr>
<td>12</td>
<td>0.1</td>
<td>$6.0 \times 10^{-8}$</td>
<td>$4.8 \times 10^{-8}$</td>
<td>$6.0 \times 10^{-8}$</td>
<td>$8.1 \times 10^{-10}$</td>
<td>$1.6 \times 10^{-10}$</td>
</tr>
<tr>
<td>12</td>
<td>0.001</td>
<td>$3.9 \times 10^{-28}$</td>
<td>$(1.8 \times 10^{-40})$</td>
<td>$3.9 \times 10^{-28}$</td>
<td>$(3.2 \times 10^{-74})$</td>
<td>$1.4 \times 10^{-55}$</td>
</tr>
</tbody>
</table>