Rare event simulation applied to dependability analysis

Bruno Tuffin
(based on joint works with
Z. Botev, P. L'Ecuyer, G. Rubino, S. Saggadi, K.S. Trivedi)

INRIA Rennes - Centre Bretagne Atlantique

Paris
2015

Outline

Outline

Introduction: rare events

Rare events occur when dealing with performance evaluation in many different areas

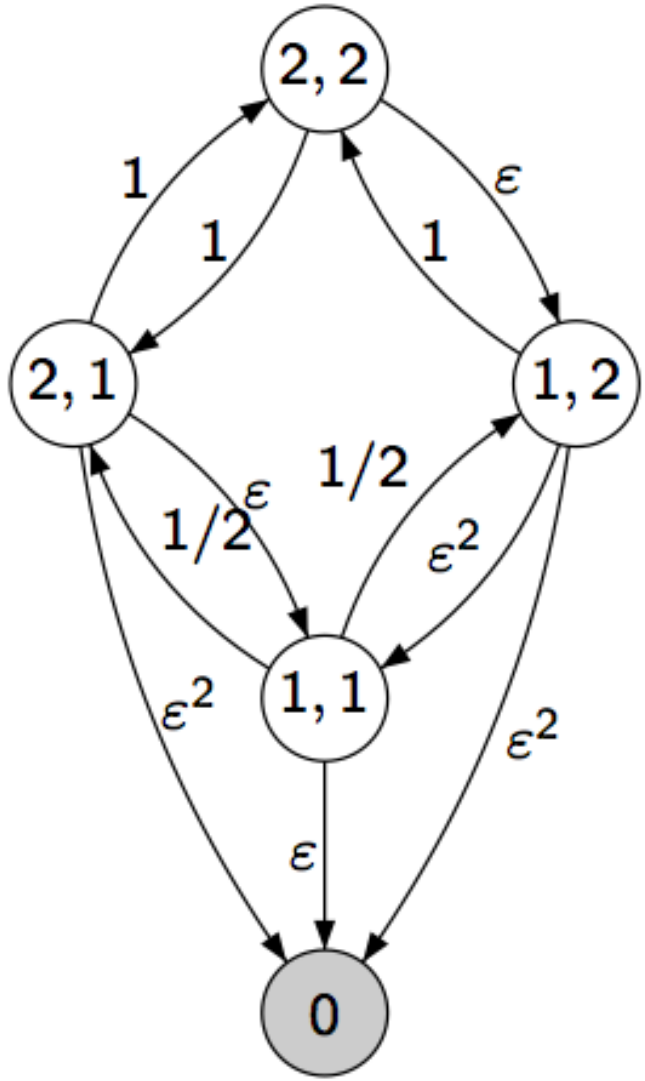
- in *telecommunication networks*: loss probability of a small unit of information (a packet, or a cell in ATM networks), connectivity of a set of nodes,
- in *dependability analysis*: probability that a system is failed at a given time, availability, mean-time-to-failure,
- in *air control systems*: 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),
- ...

What is a rare event? Why simulation?

- A rare event is an event occurring with a small probability.
- How small? Depends on the context.
- In many cases, these probabilities can be between 10^{-8} and 10^{-10}, or even at lower values. Main example: critical systems, that is,
  - systems where the rare event is a catastrophic failure with possible human losses,
  - or systems where the rare event is a catastrophic failure with possible monetary losses.
- In most of the above problems, the mathematical model is often too complicated to be solved by analytic or numeric methods because
  - the assumptions are not stringent enough,
  - the mathematical dimension of the problem is too large,
  - the state space is too large to get a result in reasonable time,
  - ...
- Simulation is, most of the time, the only tool at hand.

Example: Highly Reliable Markovian Systems (HRMS)

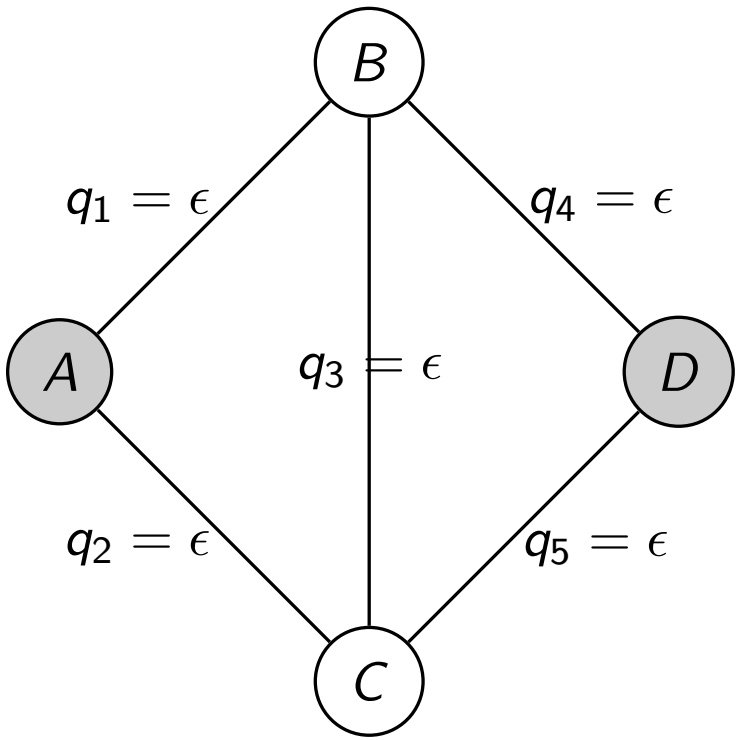
- System with c types of components. $Y = (Y_1, \ldots, Y_c)$ with Y_i number of up components.
- $\mathbf{1}$: state with all components up.
- Markov chain. Failure rates are $O(\varepsilon)$, but not repair rates. Failure propagations possible.
- System down when in grey state(s).
- Goal: compute $\mu(y)$ probability to hit Δ before $\mathbf{1}$.
- $\mu(\mathbf{1})$ important in dependability analysis,
- Small if ε small.



Example: connectivity within a graph

- *Static* reliability problems (*time* is not an explicit variable)
- Communication network:
  - nodes assumed to be perfect,
  - links can fail independently.
  - For each edge e, *elementary unreliability* q_e, reliability $r_e = 1 - q_e$.
  - The network works iff two specific nodes communicate.
- Model: graph with M links
- $X = (X_1, \ldots, X_M)$ (random) *configuration* with $X_e = 1$ if edge e works, 0 otherwise.
- state of the system: $\phi(X)$, where $\phi(X) = 1$ iff s and t not connected.

- $u = \mathbb{E}[\phi(X)]$, computation NP-hard problem in general.
- u small if individual unreliabilities small and/or redundancy of paths.



Outline

Monte Carlo

- In all the above problems, the goal is to compute $\mu = \mathbb{E}[X]$ of some random variable X.
- Monte Carlo simulation (in its basic form) generates n independent copies of X, $(X_i, 1 \leq i \leq n)$,
- $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$ approximation of μ.
- Almost sure convergence as $n \to \infty$ (law of large numbers).
- Accuracy: central limit theorem, yielding a confidence interval

$$\mu \in \left(\bar{X}_n - \frac{c_\alpha \sigma}{\sqrt{n}}, \bar{X}_n + \frac{c_\alpha \sigma}{\sqrt{n}} \right)$$

  - α: desired confidence probability,
  - $c_\alpha = \Phi^{-1}(1 - \frac{\alpha}{2})$ with Φ is the cumulative Normal distribution function of $\mathcal{N}(0, 1)$
  - $\sigma^2 = \mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$, estimated by $S_n^2 = (1/(n-1))\sum_{i=1}^{n} X_i^2 - (n/(n-1))(\bar{X}_n)^2$.

Inefficiency of crude Monte Carlo

- *Crude* Monte Carlo: simulates the model directly
- We compute the probability $\mu = \mathbb{E}[1_A] << 1$ of a rare event A.
- X_i Bernoulli r.v.: 1 if the event is hit and 0 otherwise.
- To get a single occurence, we need in average $1/\mu$ replications (10^9 for $\mu = 10^{-9}$), and more to get a confidence interval.
- In most cases, you will get $(0, 0)$ as a confidence interval.
- $n\bar{X}_n$ Binomial with parameters (n, μ) and the confidence interval is

$$\left(\bar{X}_n - \frac{c_\alpha \sqrt{\mu(1-\mu)}}{\sqrt{n}}, \bar{X}_n + \frac{c_\alpha \sqrt{\mu(1-\mu)}}{\sqrt{n}} \right).$$

- *Relative half width* $c_\alpha \sigma/(\sqrt{n}\mu) = c_\alpha \sqrt{(1-\mu)/\mu/n} \to \infty$ as $\mu \to 0$.
- For a given relative error RE , the required value of

$$n = (c_\alpha)^2 \frac{1-\mu}{RE^2 \mu},$$

  inversely proportional to μ.
- Two main families of techniques:
  - Splitting (also called *subset simulation*) and Importance Sampling.

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 *bounded relative variance* (or *bounded relative error*) if $\sigma^2(X(\epsilon))/\mu^2(\epsilon)$ is bounded uniformly in ϵ.
- 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: *asymptotic optimality* (or *logarithmic efficiency*) if $\lim_{\epsilon \to 0} \ln(\mathbb{E}[X^2(\epsilon)])/\ln(\mu(\epsilon)) = 2$.
- Stronger property: *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

Outline

Importance Sampling (IS)

- Let $X = h(Y)$ for some function h where Y obeys some probability law $\mathbb{P}$.
- IS replaces $\mathbb{P}$ by another probability measure $\tilde{\mathbb{P}}$, using

$$E[X] = \int h(y) d\mathbb{P}(y) = \int h(y) \frac{d\mathbb{P}(y)}{d\tilde{\mathbb{P}}(y)} d\tilde{\mathbb{P}}(y) = \tilde{\mathbb{E}}\left[h(Y)L(Y)\right]$$

  - $L = d\mathbb{P}/d\tilde{\mathbb{P}}$ likelihood ratio,
  - $\tilde{\mathbb{E}}$ is the expectation associated to probability law $\mathbb{P}$.
- Required condition: $d\tilde{\mathbb{P}}(y) \neq 0$ when $h(y)d\mathbb{P}(y) \neq 0$.
- If $\mathbb{P}$ and $\tilde{\mathbb{P}}$ continuous laws, L ratio of density functions.
- If $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are discrete laws, L ratio of indiv. prob.
- Unbiased estimator: $\frac{1}{n}\sum_{i=1}^{n} h(Y_i)L(Y_i)$ with $(Y_i,\, 1 \leq i \leq n)$ i.i.d; copies of Y, according to $\tilde{\mathbb{P}}$.
- Goal: select probability law $\tilde{\mathbb{P}}$ such that

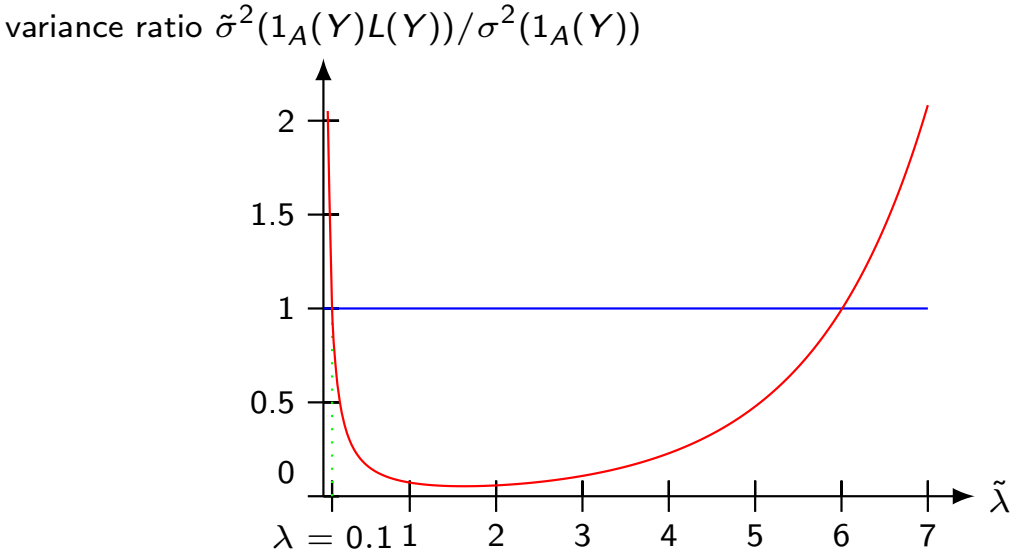
$$\tilde{\sigma}^2[h(Y)L(Y)] = \tilde{\mathbb{E}}[(h(Y)L(Y))^2] - \mu^2 < \sigma^2[h(Y)].$$

IS difficulty: system with exponential failure time

- Goal: to compute μ that the system fails before T, $\mu = \mathbb{E}[1_A(Y)] = 1 - e^{-\lambda T}$.
- Use for IS an exponential density with a different rate $\tilde{\lambda}$

$$\tilde{\mathbb{E}}[(1_A(Y)L(Y))^2] = \int_0^T \left(\frac{\lambda e^{-\lambda y}}{\tilde{\lambda} e^{-\tilde{\lambda} y}}\right)^2 \tilde{\lambda} e^{-\tilde{\lambda} y} dy = \frac{\lambda^2(1 - e^{-(2\lambda - \tilde{\lambda})T})}{\tilde{\lambda}(2\lambda - \tilde{\lambda})}.$$

- Variance ratio for $T = 1$ and $\lambda = 0.1$:



Conclusions from the example

- Increasing the failure rate ($\tilde{\lambda} > \lambda$) reduces drastically the variance
- But increasing it too much has the opposite effect!
- Trade-off (how much to modify) difficult in practice.
- To be applied at each level, but can be very efficient with a good choice.
- Often applied principle:
  - consider a parameterized family of IS change of measures
  - determine the parameter minimizing the variance.
  - Ex: what we have done with the previous example.

Optimal estimator for estimating $\mathbb{E}[h(Y)] = \int h(y)L(y)d\tilde{\mathbb{P}}(y)$

- Optimal change of measure:

$$d\tilde{\mathbb{P}} = \frac{|h(Y)|}{\mathbb{E}[|h(Y)|]} d\mathbb{P}.$$

- *Proof:* for any alternative IS measure $\mathbb{P}'$, leading to the likelihood ratio L' and expectation $\mathbb{E}'$,

$$\tilde{\mathbb{E}}[(h(Y)L(Y))^2] = (\mathbb{E}[|h(Y)|])^2 = (\mathbb{E}'[|h(Y)|L'(Y)])^2 \leq \mathbb{E}'[(h(Y)L'(Y))^2].$$

- If $h \geq 0$, $\tilde{\mathbb{E}}[(h(Y)L(Y))^2] = (\mathbb{E}[h(Y)])^2$, i.e., $\tilde{\sigma}^2(h(Y)L(Y)) = 0$. That is, IS provides a zero-variance estimator.
- Implementing it requires knowing $\mathbb{E}[|h(Y)|]$, i.e. what we want to compute; if so, no need to simulate!
- But provides a hint on the general form of a "good" IS. measure.

Examples

- If we want to compute the probability $\mathbb{P}[A]$ of an event A, this gives

$$d\tilde{\mathbb{P}}(y) = \frac{\mathbf{1}(y \in A)}{p} d\mathbb{P}(y).$$

- For the example of computing the probability that an exponential law with rate λ is smaller than T, the optimal density (yielding variance zero) is

$$\tilde{g}(y) = g(y) \frac{\mathbf{1}(y \in [0, T])}{\mathbb{E}[\mathbf{1}(y \in [0, T])]} = \mathbf{1}(y \in [0, T]) \frac{\lambda e^{-\lambda y}}{1 - e^{-\lambda y}}.$$

  It is the same exponential distribution, truncated to the interval $[0, T]$.

IS for a discrete-time Markov chain (DTMC) $\{Y_j, j \geq 0\}$

- $X = h(Y_0, \ldots, Y_\tau)$ function of the sample path with
  - $P = (P(y,z)$ transition matrix, $\pi_0(y) = \mathbb{P}[Y_0 = y]$, initial probabilities
  - up to a stopping time τ, first time it hits a set Δ.
  - $\mu(y) = \mathbb{E}_y[X]$.
- IS replaces the probabilities of paths $(y_0, \ldots, y_n)$,

$$\mathbb{P}[(Y_0, \ldots, Y_\tau) = (y_0, \ldots, y_n)] = \pi_0(y_0) \prod_{j=1}^{n-1} P(y_{j-1}, y_j),$$

  by $\tilde{\mathbb{P}}[(Y_0, \ldots, Y_\tau) = (y_0, \ldots, y_n)]$ st $\tilde{\mathbb{E}}[\tau] < \infty$.
- For convenience, the IS measure remains a DTMC, replacing $P(y,z)$ by $\tilde{P}(y,z)$ and $\pi_0(y)$ by $\tilde{\pi}_0(y)$.
- Then $L(Y_0, \ldots, Y_\tau) = \dfrac{\pi_0(Y_0)}{\tilde{\pi}_0(Y_0)} \displaystyle\prod_{j=1}^{\tau-1} \frac{P(Y_{j-1}, Y_j)}{\tilde{P}(Y_{j-1}, Y_j)}$.

Zero-variance IS estimator for Markov chains simulation

- Restrict to an additive (positive) cost

$$X = \sum_{j=1}^{\tau} c(Y_{j-1}, Y_j)$$

- Is there a Markov chain change of measure yielding zero-variance?
- Yes we have zero variance with

$$\begin{aligned} \tilde{P}(y,z) &= \frac{P(y,z)(c(y,z)+\mu(z))}{\sum_w P(y,w)(c(y,w)+\mu(w))} \\ &= \frac{P(y,z)(c(y,z)+\mu(z))}{\mu(y)}. \end{aligned}$$

- Without the additivity assumption the probabilities for the next state must depend in general of the entire history of the chain.

Zero-variance for Markov chains

- Proof by induction on the value taken by τ, using the fact that $\mu(Y_\tau) = 0$ In that case, if $\tilde{X}$ denotes the IS estimator,

$$\begin{aligned}
\tilde{X} &= \sum_{i=1}^{\tau} c(Y_{i-1}, Y_i) \prod_{j=1}^{i} \frac{P(Y_{j-1}, Y_j)}{\tilde{P}(Y_{j-1}, Y_j)} \\
&= \sum_{i=1}^{\tau} c(Y_{i-1}, Y_i) \prod_{j=1}^{i} \frac{P(Y_{j-1}, Y_j)\mu(Y_{j-1})}{P(Y_{j-1}, Y_j)(c(Y_{j-1}, Y_j) + \mu(Y_j))} \\
&= \sum_{i=1}^{\tau} c(Y_{i-1}, Y_i) \prod_{j=1}^{i} \frac{\mu(Y_{j-1})}{c(Y_{j-1}, Y_j) + \mu(Y_j)} \\
&= \mu(Y_0)
\end{aligned}$$

- *Unique* Markov chain implementation of the zero-variance estimator.
- Again, implementing it requires knowing $\mu(y)\ \forall y$, the quantities we wish to compute.
- Approximation to be used.

Zero-variance approximation

- Use a heuristic approximation $\hat{\mu}(\cdot)$ and plug it into the zero-variance change of measure instead of $\mu(\cdot)$.
- More efficient but also more requiring technique: *learn adaptively* function $\mu(\cdot)$, and still plug the approximation into the zero-variance change of measure formula instead of $\mu(\cdot)$.
  - *Adaptive Monte Carlo* (AMC) proceeds iteratively.
    - Considers several steps and n_i independent simulation replications at step i.
    - At step i, replaces $\mu(x)$ by a guess $\mu^{(i)}(x)$
    - use probabilities $\tilde{P}^{(i)}_{y,z} = \dfrac{P_{y,z}(c_{y,z} + \mu^{(i)}(z))}{\sum_w P_{y,w}(c_{y,w} + \mu^{(i)}(w))}$.
    - Gives a new estimation $\mu^{(i+1)}(y)$ of $\mu(y)$, from which a new transition matrix $\tilde{P}^{(i+1)}$ is defined.
  - *Adaptive stochastic approximation* (ASA) updates the probabilities at each step of the simulation.
  - But those methods require to store a lot of information for large systems.
- Other methods, based on subsolutions of Isaac equations (P. Dupuis *et al.*) or the construction of Lyapounov functions (Blanchet, Glynn *et al.*).

Illustration of heuristics: birth-death process

- Let $P(i, i+1) = p$ and $P(i, i-1) = 1 - p$ for $1 \leq i \leq B-1$, and $P(0,1) = P(B, B-1) = 1$.
- We want to compute $\mu(1)$, probability of reaching B before coming back to 0.
- If p small, to approach $\mu(\cdot)$, we can use

$$\hat{\mu}(y) = p^{B-y} \quad \forall y \in \{1, \ldots, B-1\}$$

  with $\hat{\mu}(0) = 0$ and $\hat{\mu}(B) = 1$ based on the asymptotic estimate $\mu(i) = p^{B-i} + o(p^{B-i})$.
- We can verify that the variance of this estimator is going to 0 (for fixed sample size) as $p \to 0$.

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.
- Failure rates are $O(\varepsilon)$, but not repair rates. Failure propagations possible.
- System down (in Δ) when some combinations of components are down.
- Goal: compute $\mu(\mathbf{1})$ with $\mu(y)$ probability to hit Δ before $\mathbf{1}$.
- Simulation using the embedded DTMC. Failure probabilities are $O(\varepsilon)$ (except from $\mathbf{1}$). How to improve (accelerate) this?
- Existing method: $\forall y \neq \mathbf{1}$, 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

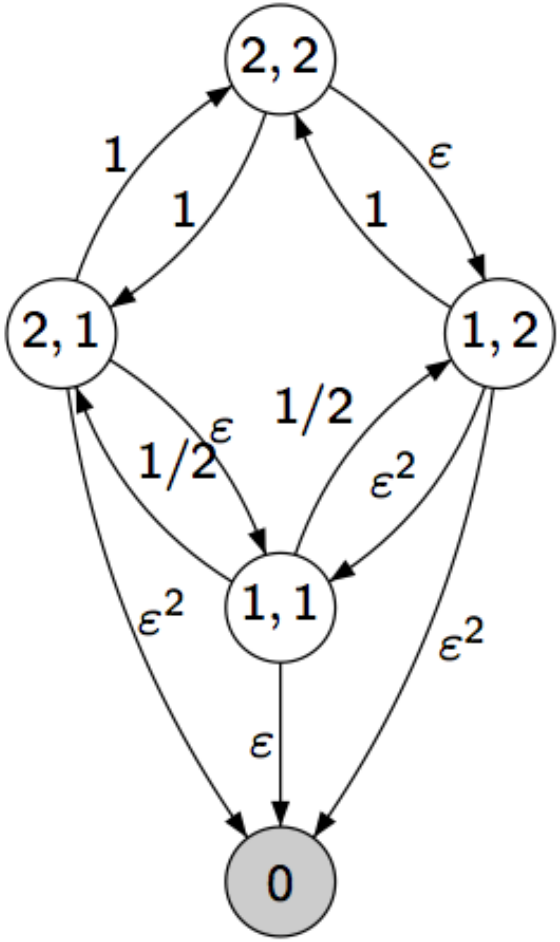


Figure: Original probabilities

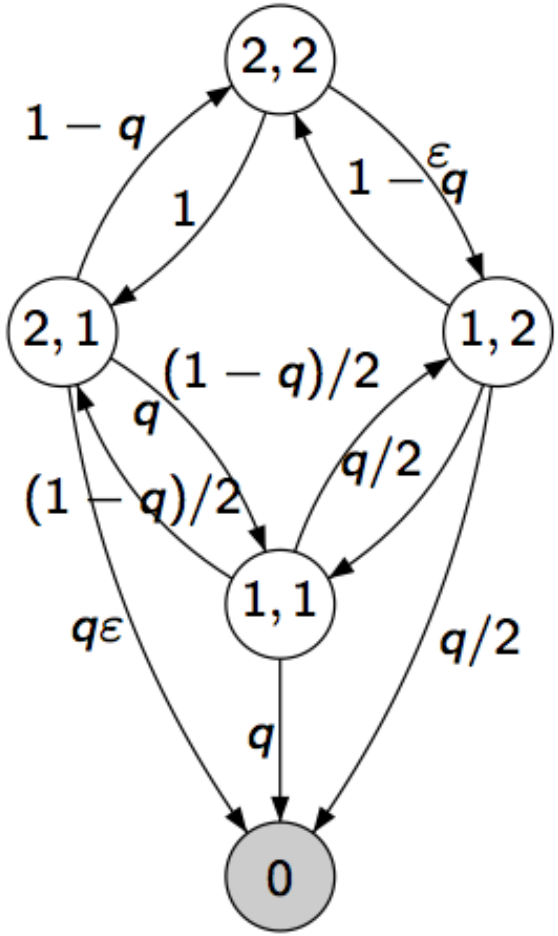


Figure: Probabilities under IS/BFB

HRMS, Zero-variance IS

L'Ecuyer & T., ANOR, 2011

- 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 Δ 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 ϵ.

- 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.

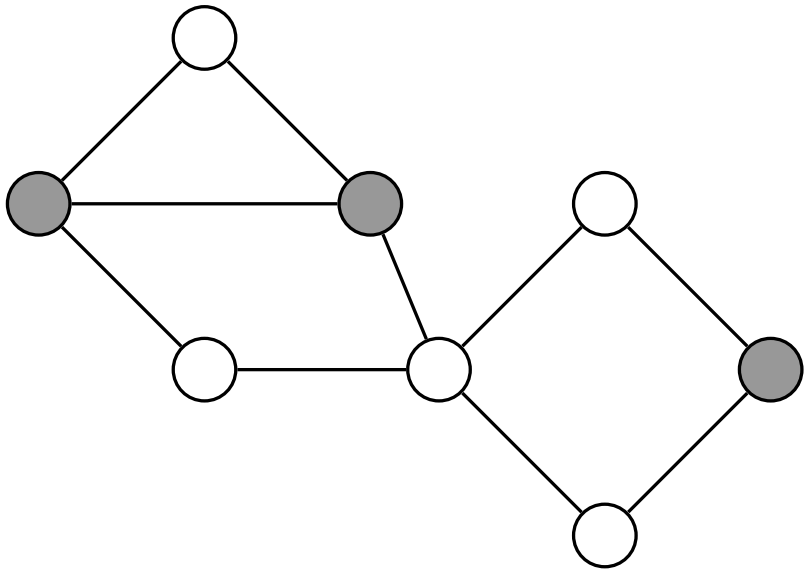
HRMS: numerical illustrations

- 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	ε	μ_0	BFB est	ZVA est	BFB σ^2	ZVA σ^2
3	0.001	2.6×10^{-3}	2.7×10^{-3}	2.6×10^{-3}	6.2×10^{-5}	2.2×10^{-8}
6	0.01	1.8×10^{-7}	1.9×10^{-7}	1.8×10^{-7}	6.3×10^{-11}	2.0×10^{-14}
6	0.001	1.7×10^{-11}	1.8×10^{-11}	1.7×10^{-11}	8.8×10^{-19}	1.2×10^{-23}
12	0.1	6.0×10^{-8}	4.8×10^{-8}	6.0×10^{-8}	8.1×10^{-10}	1.6×10^{-10}
12	0.001	3.9×10^{-28}	(1.8×10^{-40})	3.9×10^{-28}	(3.2×10^{-74})	1.4×10^{-55}

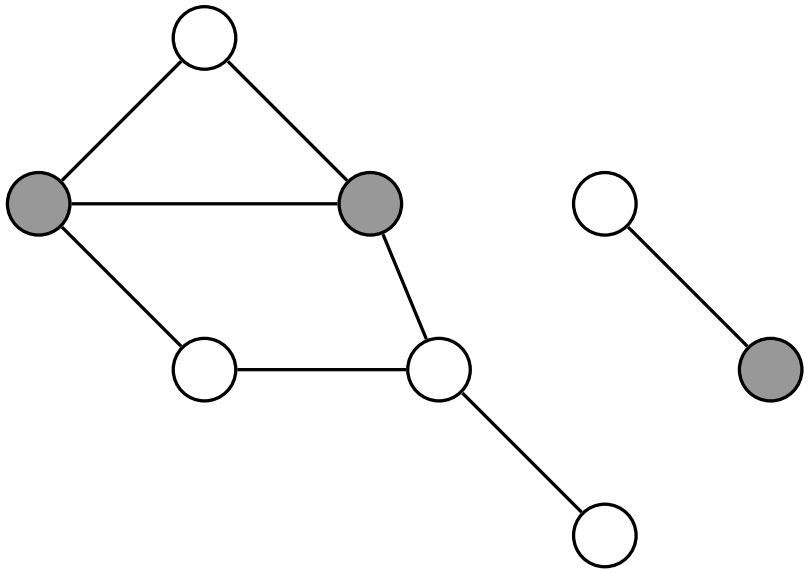
Graph model

- 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 $\mathcal{G}$)?
- $X = (X_1, \ldots, X_M)$ (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.
- $u = \mathbb{E}[\phi(X)] = \sum_{x \in \{0,1\}^M} \phi(x) \mathbb{P}[X = x]$.



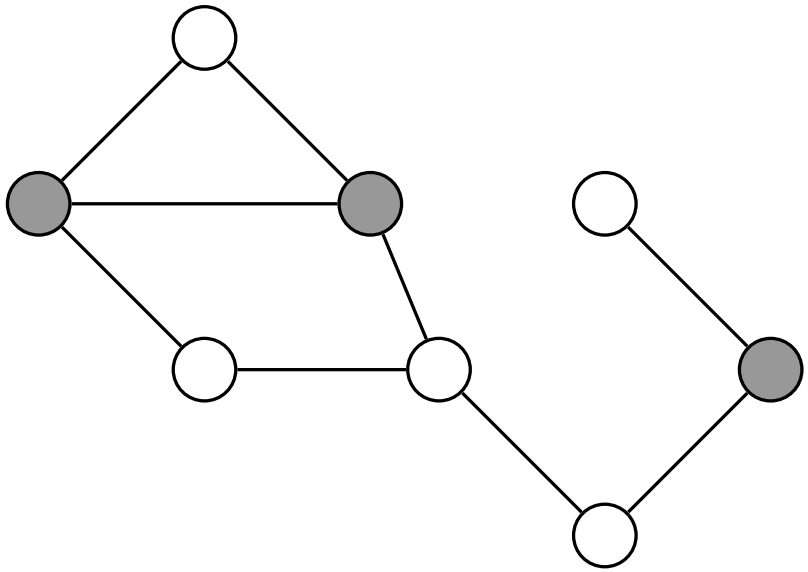
Graph model

- 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 $\mathcal{G}$)?
- $X = (X_1, \ldots, X_M)$ (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.
- $u = \mathbb{E}[\phi(X)] = \sum_{x \in \{0,1\}^M} \phi(x) \mathbb{P}[X = x]$.



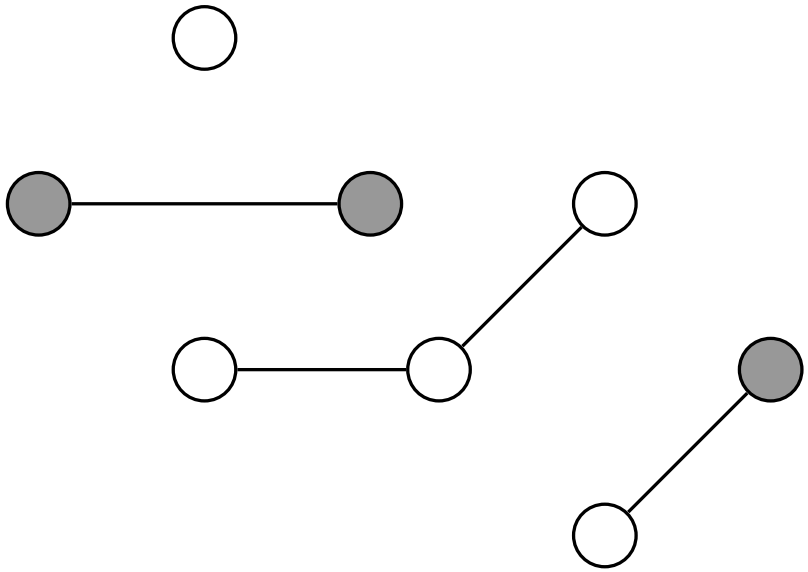
Graph model

- 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 $\mathcal{G}$)?
- $X = (X_1, \ldots, X_M)$ (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.
- $u = \mathbb{E}[\phi(X)] = \sum_{x \in \{0,1\}^M} \phi(x) \mathbb{P}[X = x]$.



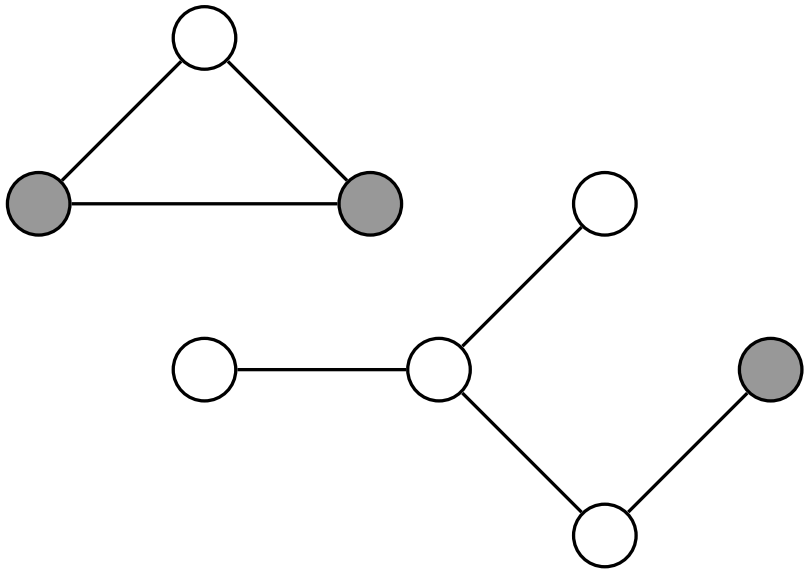
Graph model

- 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 $\mathcal{G}$)?
- $X = (X_1, \ldots, X_M)$ (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.
- $u = \mathbb{E}[\phi(X)] = \sum_{x \in \{0,1\}^M} \phi(x) \mathbb{P}[X = x]$.



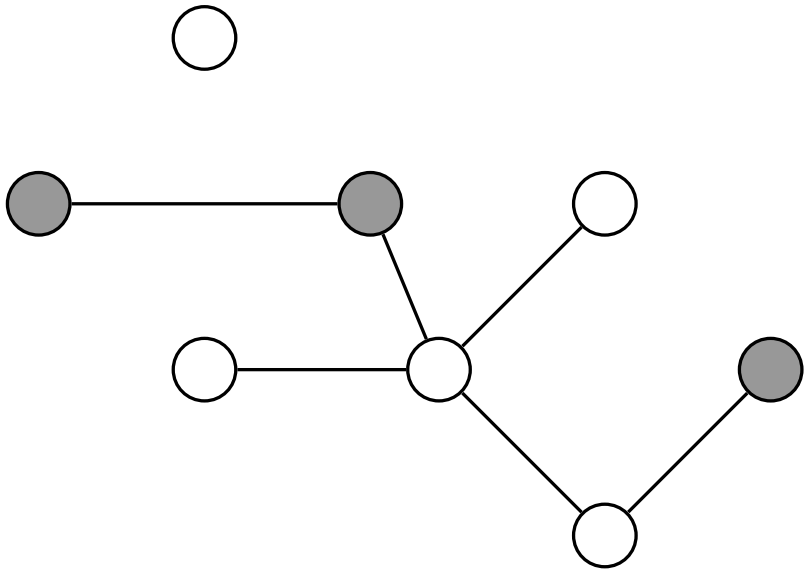
Graph model

- 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 $\mathcal{G}$)?
- $X = (X_1, \ldots, X_M)$ (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.
- $u = \mathbb{E}[\phi(X)] = \sum_{x \in \{0,1\}^M} \phi(x) \mathbb{P}[X = x]$.



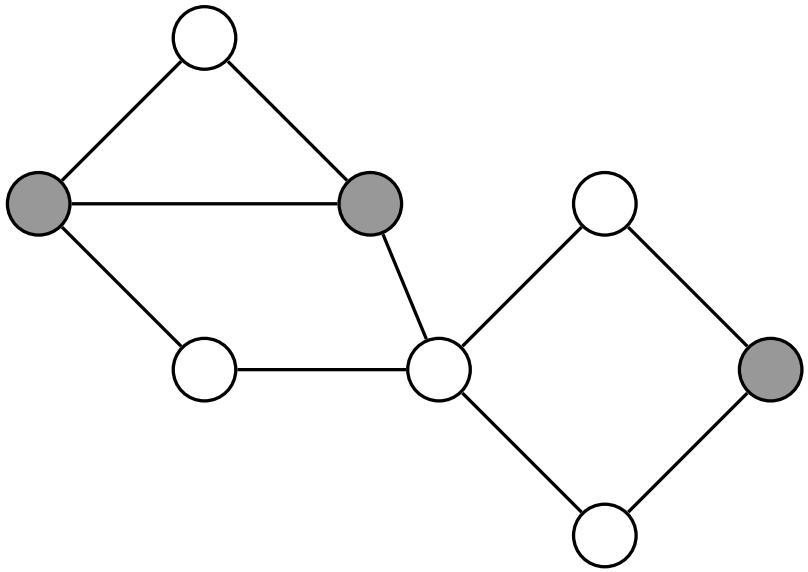
Graph model

- 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 $\mathcal{G}$)?
- $X = (X_1, \ldots, X_M)$ (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 $\mathcal{G}$)?
- $X = (X_1, \ldots, X_M)$ (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.
- $u = \mathbb{E}[\phi(X)] = \sum_{x \in \{0,1\}^M} \phi(x) \mathbb{P}[X = x]$.



- We have to sum over the 2^M configurations.

Zero-variance est. L'Ecuyer, Rubino, Saggadi & T., IEEE Trans. Rel. 2011

- 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:

$$r'_m(x_1, \cdots, x_{m-1}) = \frac{r_m u_{m+1}(x_1, \cdots, x_{m-1}, 1)}{r_m u_{m+1}(x_1, \cdots, x_{m-1}, 1) + (1 - r_m) u_{m+1}(x_1, \cdots, x_{m-1}, 0)}.$$

- Remark (by conditionning) that

$$u_m(x_1, \cdots, x_{m-1}) = r_m u_{m+1}(x_1, \cdots, x_{m-1}, 1) + (1 - r_m) r_m u_{m+1}(x_1, \cdots, x_{m-1}, 0).$$

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

$$r'_m(x_1, \cdots, x_{m-1}) = \frac{r_m \hat{u}_{m+1}(x_1, \cdots, x_{m-1}, 1)}{r_m \hat{u}_{m+1}(x_1, \cdots, x_{m-1}, 1) + (1 - r_m)\hat{u}_{m+1}(x_1, \cdots, x_{m-1}, 0)}.$$

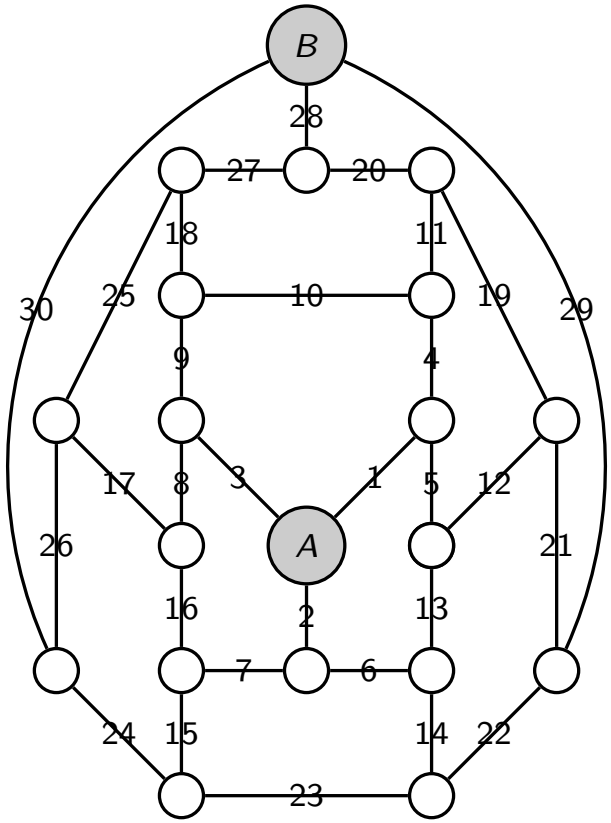
Approximation of the zero-variance estimator

- 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 $\mathcal{K}$-cut) is a set of edges such that, if we remove them, the nodes in $\mathcal{K}$ 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.
- In a given state $(x_1, \cdots, x_{m-1})$, we need to determine $\hat{u}_{m+1}(x_1, \cdots, x_{m-1}, 1)$ and $\hat{u}_{m+1}(x_1, \cdots, x_{m-1}, 0)$.
- This adds some computational burden, but should substantially reduce the variance.

Proposition

Bounded relative error proved in general,
Vanishing relative error under identified conditions.

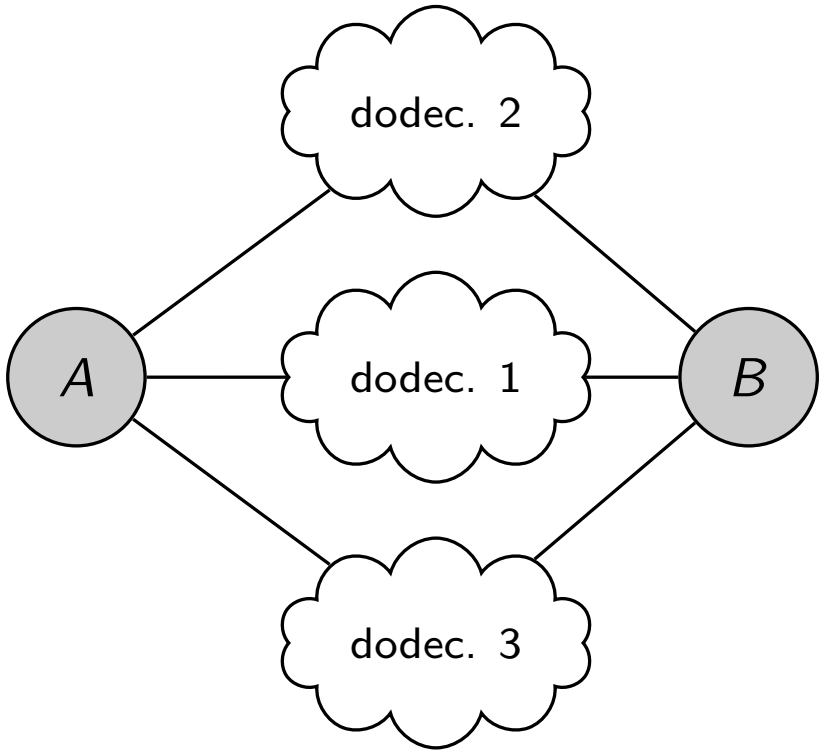
Ex: dodecahedron topology, all links with unreliability ϵ



ϵ	Estimation	Confidence interval	Std deviation	Relative error
10^{-1}	$2.8960\ 10^{-3}$	$(2.8276\ 10^{-3}, 2.9645\ 10^{-3})$	$3.49\ 10^{-3}$	1.2
10^{-2}	$2.0678\ 10^{-6}$	$(2.0611\ 10^{-6}, 2.0744\ 10^{-6})$	$3.42\ 10^{-7}$	0.17
10^{-3}	$2.0076\ 10^{-9}$	$(2.0053\ 10^{-9}, 2.0099\ 10^{-9})$	$1.14\ 10^{-10}$	0.057
10^{-4}	2.0007^{-12}	$(2.0000\ 10^{-12}, 2.0014\ 10^{-12})$	$3.46\ 10^{-14}$	0.017

With respect to crude MC, a computational time increase of 16.

Larger networks: 3 dodecahedrons in parallel



ϵ	Estimate	95% confidence interval	std dev.	Relative Error
10^{-1}	2.3573×10^{-8}	$(2.2496 \times 10^{-8},\ 2.4650 \times 10^{-8})$	5.49×10^{-8}	2.3
5×10^{-2}	2.5732×10^{-11}	$(2.5138 \times 10^{-11},\ 2.6327 \times 10^{-11})$	3.03×10^{-11}	1.2
10^{-2}	8.7655×10^{-18}	$(8.7145 \times 10^{-18},\ 8.8165 \times 10^{-18})$	2.60×10^{-18}	0.30

- 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.

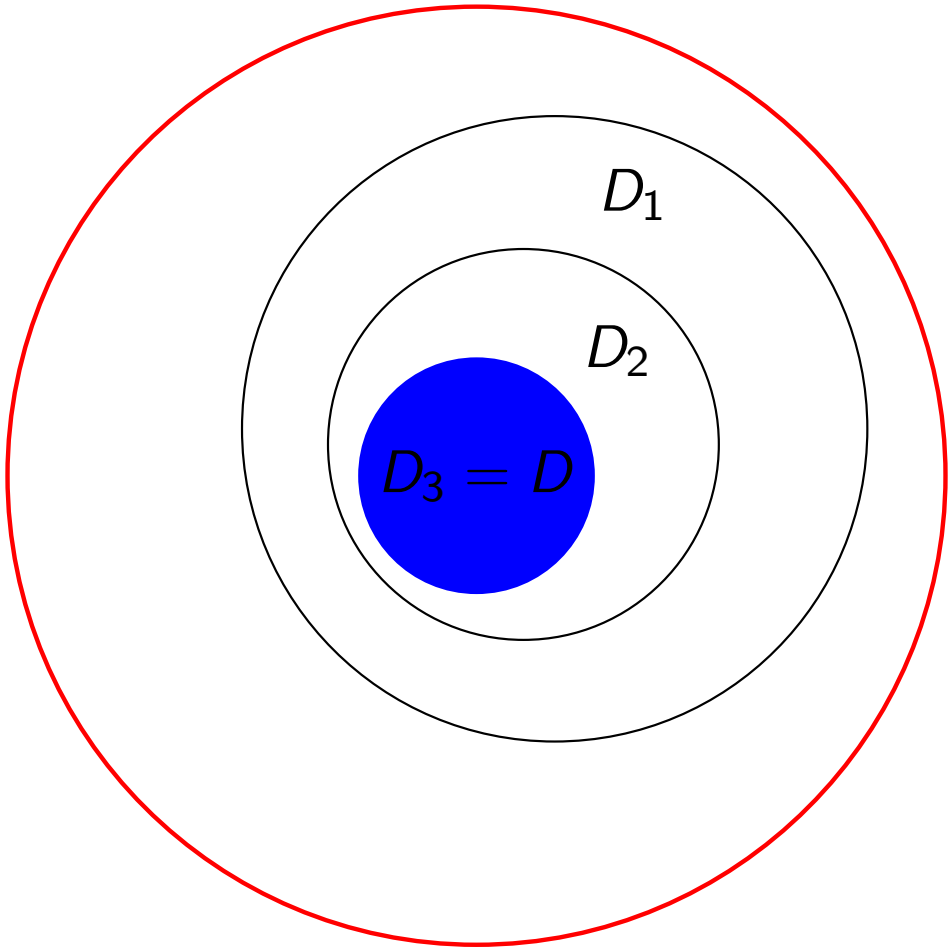
Outline

Splitting: general principle

- Splitting is the other main rare event simulation technique.
- Assume we want to compute the probability $\mathbb{P}(D)$ of an event D.
- General idea:
  - Decompose
$$D_1 \supset \cdots \supset D_m = D,$$
  - Use $\mathbb{P}(D) = \mathbb{P}(D_1)\mathbb{P}(D_2 \mid D_1) \cdots \mathbb{P}(D_m \mid D_{m-1})$, each conditional event being "not rare",
  - Estimate each individual conditional probability by crude Monte Carlo, i.e., without changing the laws driving the model.
  - The final estimate is the product of individual estimates.
- Question: how to do it for a stochastic process? Difficult to sample conditionally to an intermediate event.

Rem: Generalized Splitting not presented here (cf papers with P. L'Ecuyer and Z. Botev). Makes use of Gibbs sampling.

Graphical interpretation



Splitting and Markov chain $\{Y_j; j \geq 0\} \in \mathcal{Y}$

- Goal: compute $\gamma_0 = \mathbb{P}[\tau_B < \tau_A]$ with
  - $\tau_A = \inf\{j > 0 : Y_{j-1} \notin A \text{ and } Y_j \in A\}$
  - $\tau_B = \inf\{j > 0 : Y_j \in B\}$
- Intermediate levels from importance function $h : \mathcal{Y} \to \mathbb{R}$ with $A = \{x \in \mathcal{Y} : h(x) \leq 0\}$ and $B = \{x \in \mathcal{Y} : h(x) \geq \ell\}$:
  - Partition $[0, \ell)$ in m subintervals with boundaries $0 = \ell_0 < \ell_1 < \cdots < \ell_m = \ell$.
  - Let $T_k = \inf\{j > 0 : h(Y_j) \geq \ell_k\}$ and $D_k = \{T_k < \tau_A\}$.
- 1st stage:
  - simulate N_0 chains until $\min(\tau_A, T_1)$.
  - If R_1 number of chains for which D_1 occurs, $\hat{p}_1 = R_1/N_0$ unbiased estimator of $p_1 = \mathbb{P}(D_1)$.
- Stage $1 < k \leq m$:
  - If $R_{k-1} = 0$, $\hat{p}_l = 0$ for all $l \geq k$ and the algorithm stops
  - Otherwise, start N_k chains from these R_k entrance states, by potentially cloning (splitting) some chains
  - simulate these chains up to $\min(\tau_A, T_k)$.
  - $\hat{p}_k = R_k/N_{k-1}$ unbiased estimator of $p_k = \mathbb{P}(D_k | D_{k-1})$

The different implementations

- *Fixed splitting*:
  - clone each of the R_k chains reaching level k in c_k copies, for a fixed positive integer c_k.
  - $N_k = c_k R_k$ is random.
- *Fixed effort*:
  - N_k fixed a priori
  - *random assignment* draws the N_k starting states at random, with replacement, from the R_k available states.
  - *fixed assignment*, on the other hand, we would split each of the R_k states approximately the same number of times.
  - Fixed assignment gives a smaller variance than random assignment because it amounts to using stratified sampling over the empirical distribution G_k at level k.
- Fixed splitting can be implemented in a depth-first way, recursively, while fixed effort cannot.
- On the other hand, you have no randomness (less variance) in the number of chains with fixed effort.

Diminishing the computational effort

- As k increases, it is likely that the average time before reaching the next level or going back to A increases significantly.
- We can kill (truncate) trajectories hat go a given number β of levels down (unlikely to come back), but biased.
- Unbiased solution: apply the Russian roulette principle
  - kill the trajectory going down with a probability r_β. If it survives, assign a multiplicative weight $1/(1-r_\beta)$.
  - Several possible implementations to reduce the variance due to the introduction of weights.

Issues to be solved

- *How to define the importance function h?*
  - If the state space is one-dimensional and included in $\mathbb{R}$, the final time is an almost surely finite stopping time and the critical region is $B = [b, \infty)$, any strictly increasing function would be good (otherwise a mapping can be constructed, by just moving the levels), such as for instance $h(x) = x$.
  - If the state space is multidimensional: the importance function is a one-dimensional projection of the state space.
  - Desirable property: the probability to reach the next level should be the same, whatever the entrance state in the current level.
  - Ideally, $h(x) = \mathbb{P}[\tau_B \leq \tau_A \mid X(0) = x]$, but as in IS, they are a probabilities we are looking for.
  - This $h(\cdot)$ can also be learnt or estimated *a priori*, with a presimulation, by partitionning the state space and assuming it constant on each region.

Issues to be solved (2)

- *How many offsprings at each level?*
  - In fixed splitting:
    - ★ if $c_k < 1/p_k$, we do not split enough, it will become unlikely to reach the next event;
    - ★ if $c_k > 1/p_k$, the number of trajectories will exponentially explode with the number of levels.
    - ★ The right amount is $c_k = 1/p_k$ (c_k can be randomized to reach that value if not an integer).
  - In fixed effort, no explosion is possible.
  - In both cases, the right amount has to be found.
- *How many levels to define?*
  - i.e., what probability to reach the next level?

Optimal values

- In a general setting, very few results exist:
  - We only have a central limit theorem based on genetic type interacting particle systems, as the sample increases.
  - Nothing exist on the definition of optimal number of levels...
- Consider the simplified setting, with a single entrance state at each level.
- Similar to coin–flipping to see if next level is reached or not.
- In that case, asymptotically optimal results can be derived, providing hints of values to be used.

Simplified setting and fixed effort

- $N_0 = N_1 = \cdots = N_{m-1} = n$
- The $\hat{p}_k$'s binomial r.v. with parameters n and $p_k = p = \mu_0^{1/m}$ assumed independent.
- It can be shown that

$$\begin{aligned} \mathrm{Var}[\hat{p}_1 \cdots \hat{p}_m] &= \prod_{k=1}^{m} \mathbb{E}[\hat{p}_k^2] - \gamma_0^2 = \left(p^2 + \frac{p(1-p)}{n}\right)^m - p^{2m} \\ &= \frac{mp^{2m-1}(1-p)}{n} + \cdots + \frac{(p(1-p))^m}{n^m}. \end{aligned}$$

- Assuming $n \gg (m-1)(1-p)/p$,
  $\mathrm{Var}[\hat{p}_1 \cdots \hat{p}_m] \approx mp^{2m-1}(1-p)/n \approx m\gamma_0^{2-1/m}/n$.
- The work normalized variance $\approx [\gamma_0^n m^2]/n = \gamma_0^{2-1/m} m^2$
- Minimized at $m = -\ln(\gamma_0)/2$
- This gives $p^m = \gamma_0 = e^{-2m}$, so $p = e^{-2}$.
- But the relative error and its work-normalized version both increase toward infinity at a logarithmic rate.
- There is no asymptotic optimality either.

Simplified setting: fixed splitting

- $N_0 = n$, $p_k = p = \gamma_0^{1/m}$ for all k, and $c = 1/p$; i.e., $N_k = R_k/p$.
- The process $\{N_k,\ k \geq 1\}$ is a *branching process*.
- From standard branching process theory

$$\mathrm{Var}[\hat{p}_1 \cdots \hat{p}_m] = m(1-p)p^{2m-1}/n.$$

- If p fixed and $m \to \infty$, the squared relative error $m(1-p)/(np)$ is unbounded,
- But it is asymptotically efficient:

$$\lim_{\gamma_0 \to 0^+} \frac{\log(\mathbb{E}[\tilde{\gamma}_n^2])}{\log \gamma_0} = \lim_{\gamma_0 \to 0^+} \frac{\log(m(1-p)\gamma_0^2/(np) + \gamma_0^2)}{\log \gamma_0} = 2.$$

- Fixed splitting is asymptotically better, but it is more sensitive to the values used.

Illustrative simple example: a tandem queue

- Illustrative of the impact of the importance function.
- Two queues in tandem
  - arrival rate at the first queue is $\lambda = 1$
  - mean service time is $\rho_1 = 1/4$, $\rho_2 = 1/2$.
  - Embedded DTMC: $Y = (Y_j, j \geq 0)$ with $Y_j = (Y_{1,j}, Y_{2,j})$ number of customers in each queue after the jth event
  - $B = \{(x_1, x_2) : x_2 \geq L = 30\}$, $A = \{(0, 0)\}$.
- Goal: impact of the choice of the importance function?
- Importance functions:

$$\begin{aligned} h_1(x_1, x_2) &= x_2, \\ h_2(x_1, x_2) &= (x_2 + \min(0, x_2 + x_1 - L))/2, \\ h_3(x_1, x_2) &= x_2 + \min(x_1, L - x_2 - 1) \times (1 - x_2/L). \end{aligned}$$

Illustration, fixed effort: a tandem queue (2)

- V_N: *variance per chain*, (N times the variance of the estimator) and the *work-normalized variance per chain*, $W_N = S_N V_N$, where S_N is the expected total number of simulated steps of the N Markov chains.
- With h_1, $\hat{V}_N$ and $\hat{W}_N$ were significantly higher than for h_2 and h_3.
- Estimators rescaled as $\tilde{V}_N = 10^{18} \times \hat{V}_N$ and $\tilde{W}_N = 10^{15} \times \hat{W}_N$.

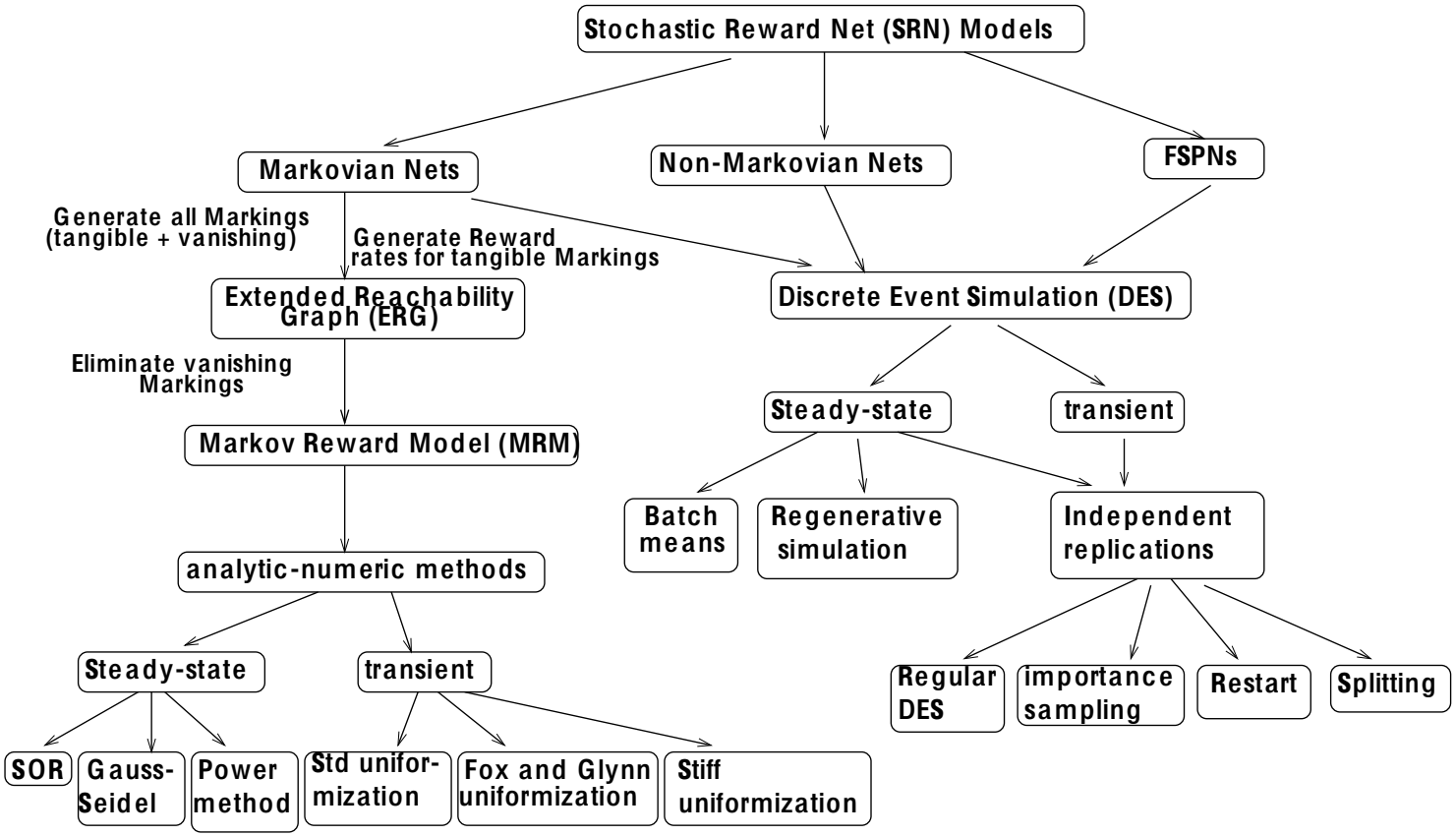
	$N = 2^{10}$		$N = 2^{12}$		$N = 2^{14}$		$N = 2^{16}$	
	$\tilde{V}_N$	$\tilde{W}_N$	$\tilde{V}_N$	$\tilde{W}_N$	$\tilde{V}_N$	$\tilde{W}_N$	$\tilde{V}_N$	$\tilde{W}_N$
h_2, Splitting	109	120	89	98	124	137	113	125
h_2, Rus. Roul.	178	67	99	37	119	45	123	47
h_3, Splitting	93	103	110	121	93	102	107	118
h_3, Rus. Roul.	90	34	93	35	94	36	109	41

Splitting and HRMS

- There are situations where the splitting method is not appropriate
- HRMS: if we want to apply splitting, the thresholds must be defined in terms of the vector of failed components (the state of the system).
- But whenever there are failed components, the next event is a repair with a very high probability.
- Regardless of how we determine the thresholds, the probabilities p_k of reaching the next threshold from the current one are always very small.
- For this reason, the splitting method cannot be made efficient in this case.

Splitting and Stochastic Petri Nets

- Work done in 1999 while at Duke University
- Basic IS implemented too.
- Implementation in SPNP (Stochastic Petri Net Package)
  - Written in CSPL (C-based SPN language), but GUI available



Implementation in SPNP

- Importance function: on a measure $P(\#p \geq x_i)$.
- Number of thresholds and threshold values can be given or obtained through a pre-simulation
- Running all trials at level i before running the trials at level $i+1 \Rightarrow$ storage requirement growing exponentially
  $\Rightarrow$ structure of the elements of the model replicated at each splitting point. The "child" paths are simulated one after another, then the structure is removed. The cloning structure needs to contain the
  1. splitting level;
  2. current weight of the path;
  3. clock;
  4. marking;
  5. list of enabled transitions and their clocks resampled conditionally.

Outline

A summary of best existing methods for static reliability estimation on the dodecahedron

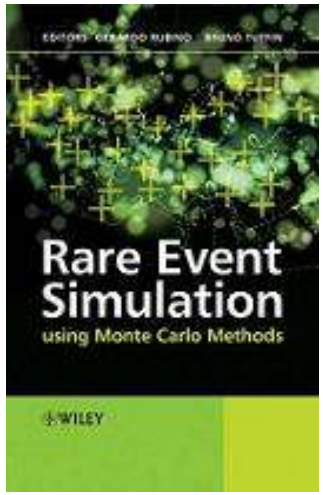
Without presenting all implementations.

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

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

Conclusions: book advertisements

Released in March 2009 (John Wiley & Sons):



In March 2010 (éditions Hermès):