UNIVERSITY OF TWENTE.

Some Advances in Importance Sampling of Reliability Models Based on Zero Variance Approximation

Daniël Reijsbergen[1] Pieter-Tjerk de Boer[1]
Werner Scheinhardt[1] Sandeep Juneja[2]

[1]University of Twente
[2]Tata Institute of Fundamental Research

RESIM workshop, June 26, 2012

Outline

1. Zero Variance Approximation (ZVA)
2. Our ZVA Algorithm
3. Variance Reduction for Free
4. Case Study

Outline

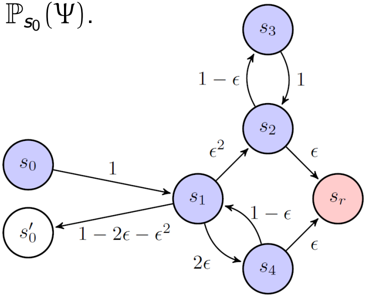
Reliability Models

DTMC with fixed state space.

Transition rates depend on rarity parameter ϵ.

Let Ψ be the event that we hit a rare state before some tabu state.

We are interested in estimating $\mathbb{P}_{s_0}(\Psi)$.



Reliability Models

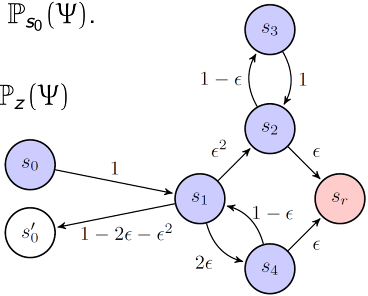
DTMC with fixed state space.

Transition rates depend on rarity parameter ϵ.

Let Ψ be the event that we hit a rare state before some tabu state.

We are interested in estimating $\mathbb{P}_{s_0}(\Psi)$.

Known: $\mathbb{P}_s(\Psi) = \sum_z \mathbb{P}(s \to z)\mathbb{P}_z(\Psi)$



Importance Sampling

State space typically large $\rightarrow$ simulation.

(Standard) Monte Carlo: generate sample paths $\omega_i, i = 1, \ldots, N$ under $\mathbb{P}$, we then estimate $\mathbb{P}(\Psi) = \mathbb{E}_{\mathbb{P}}(\mathbf{1}_\Psi)$ with

$$\frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_\Psi(\omega_i).$$

Problem: if $\mathbb{P}(\Psi)$ is small we need many simulations.

Importance sampling: generate $\omega_i, i = 1, \ldots, N$ under $\mathbb{Q}$, we then estimate $\mathbb{P}(\Psi) = \mathbb{E}_{\mathbb{P}}(\mathbf{1}_\Psi) = \mathbb{E}_{\mathbb{Q}}(L \cdot \mathbf{1}_\Psi)$ with

$$\frac{1}{N} \sum_{i=1}^{N} L(\omega_i)\mathbf{1}_\Psi(\omega_i), \quad L(\omega) = \frac{d\mathbb{P}}{d\mathbb{Q}}(\omega).$$

Zero Variance Measure

Challenge: find $\mathbb{Q}$ such that variance is lower than with $\mathbb{P}$.

Ideal: use *zero variance* measure, given by

$$\mathbb{Q}(s \to z) = \mathbb{P}(s \to z)\frac{\mathbb{P}_z(\Psi)}{\mathbb{P}_s(\Psi)} \tag{1}$$

Of course, we do not know $\mathbb{P}_{(\cdot)}(\Psi)$.

So use guess $w(s)$ for $\mathbb{P}_s(\Psi)$ to plug into (1).

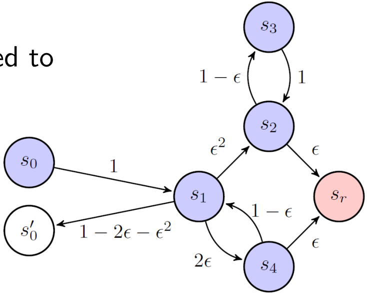
Zero Variance Approximation

How to choose good guess $w(s)$ for $\mathbb{P}_s(\Psi)$?

Define $d(s)$ = "number of ϵ's needed to reach rare from s".

Define Δ = "dominant paths": paths of order $d(s)$ to rare.

Let $w(s) = \mathbb{P}_s(\Delta)$.



Theorem: no high-probability cycles $\Rightarrow$ bounded relative error.[1]

Theorem: $\lim_{\epsilon \downarrow 0} \mathbb{P}_s(\Psi)/w(s) = 1 \Rightarrow$ vanishing relative error.

[1] P. L'Ecuyer and B. Tuffin. *Approximating zero-variance importance sampling in a reliability setting*, 2011.

Outline

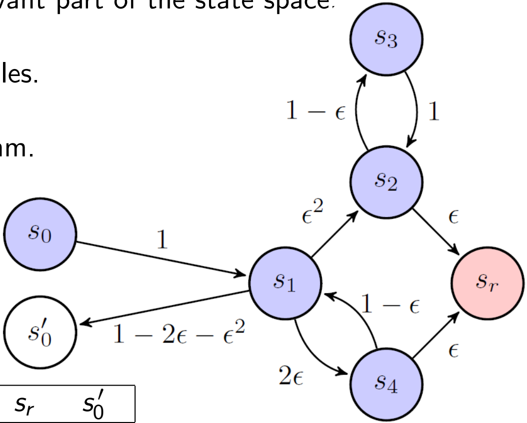
1. Zero Variance Approximation (ZVA)
2. Our ZVA Algorithm
3. Variance Reduction for Free
4. Case Study

Algorithm Sketch

Determine d and w for relevant part of the state space.

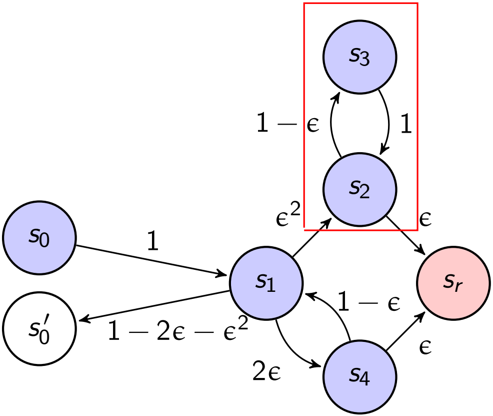
Remove high-probability cycles.

Idea: Dijkstra-based algorithm.

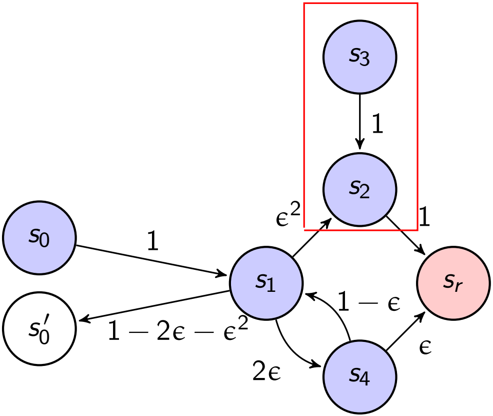


	s_0	s_1	s_2	s_3	s_4	s_r	s_0'
d'	0	0	2	2	1	2	$-\infty$
d	2	2	0	0	1	0	∞
w	$3\epsilon^2$	$3\epsilon^2$	1	1	ϵ	1	0

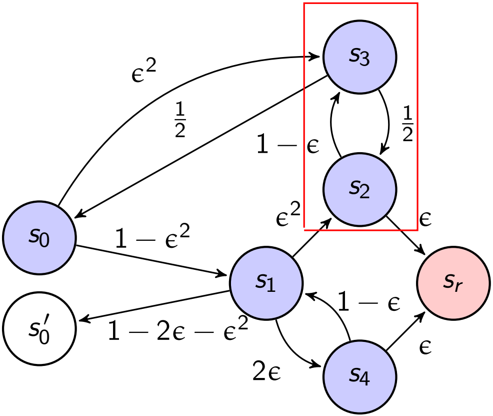
Loop Removal



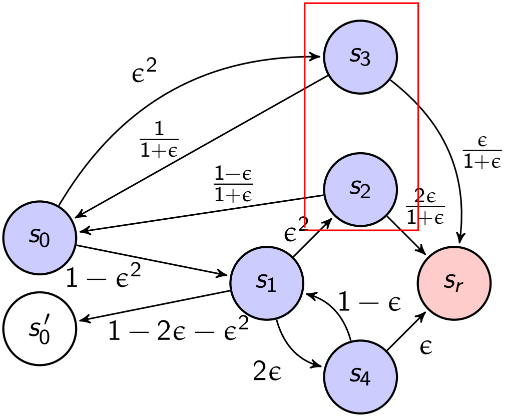
Loop Removal (cont'd)



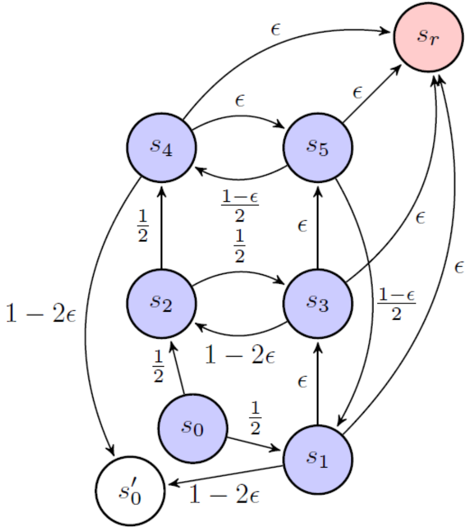
Loop Removal (cont'd)



Loop Removal (cont'd)



Loop Removal (cont'd)



Outline

Variance Reduction for Free

Idea: use knowledge of Δ, event that path is 'dominant'. [2]

Using only knowledge of $\mathbb{P}(\Delta)$ $(+)$:

$$\begin{aligned}\mathbb{P}(\Psi) &= \mathbb{P}(\Delta) + \mathbb{P}(\Psi \cap \neg\Delta) \\ &= \mathbb{P}(\Delta) + \mathbb{E}_{\mathbb{Q}}(L \cdot \mathbf{1}_{\Psi} \cdot \mathbf{1}_{\neg\Delta})\end{aligned}$$

Using knowledge of $\mathbb{P}(\Delta)$ and $\mathbb{Q}(\Delta)$ $(++)$:

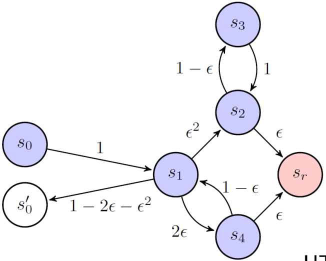
$$\mathbb{P}(\Psi) = \mathbb{P}(\Delta) + \mathbb{E}_{\mathbb{Q}}(L \cdot \mathbf{1}_{\Psi} | \neg\Delta)\mathbb{Q}(\neg\Delta)$$

[2] S. Juneja. *Estimating tail probabilities of heavy tailed distributions with asymptotically zero relative error*. 2007.

Outline

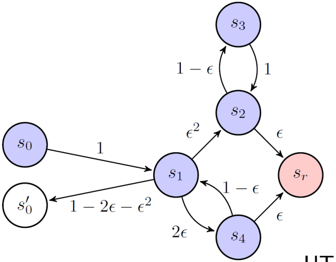
Case Study, $\epsilon = 0.1$, h = 95%-C.I. half-width

method	$\hat{p}_{\mathbb{Q}}$	$-\log(h)$	N	M
MC	$3.655 \cdot 10^{-2} \pm 1.21 \cdot 10^{-4}$	5.709	9 216 351	8 703 936
BFB	$3.654 \cdot 10^{-2} \pm 6.12 \cdot 10^{-5}$	6.392	3 600 591	2 223 172
BFB+	$3.66 \cdot 10^{-2} \pm 3.58 \cdot 10^{-5}$	6.93	3 659 717	2 287 933
BFB++	$3.657 \cdot 10^{-2} \pm 2.78 \cdot 10^{-5}$	6.946	3 678 719	2 297 880
ZVA	$3.659 \cdot 10^{-2} \pm 8.02 \cdot 10^{-6}$	8.425	2 281 743	323 533
ZVA+	$3.66 \cdot 10^{-2} \pm 2.55 \cdot 10^{-5}$	7.268	1 603 733	227 632
ZVA++	$3.659 \cdot 10^{-2} \pm 1.89 \cdot 10^{-6}$	8.892	1 340 118	189 533



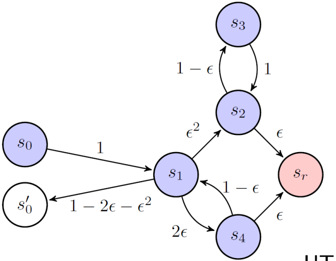
Case Study, $\epsilon = 0.1$, h = 95%-C.I. half-width

method	$\hat{p}_{\mathbb{Q}}$	$-\log(h)$	N	M
MC	$3.655{\cdot}10^{-2} \pm 1.21{\cdot}10^{-4}$	5.709	9 216 351	8 703 936
BFB	$3.654{\cdot}10^{-2} \pm 6.12{\cdot}10^{-5}$	6.392	3 600 591	2 223 172
BFB+	$3.66{\cdot}10^{-2} \pm 3.58{\cdot}10^{-5}$	6.93	3 659 717	2 287 933
BFB++	$3.657{\cdot}10^{-2} \pm 2.78{\cdot}10^{-5}$	6.946	3 678 719	2 297 880
ZVA	$3.659{\cdot}10^{-2} \pm 8.02{\cdot}10^{-6}$	8.425	2 281 743	323 533
ZVA+	$3.66{\cdot}10^{-2} \pm 2.55{\cdot}10^{-5}$	7.268	1 603 733	227 632
ZVA++	$3.659{\cdot}10^{-2} \pm 1.89{\cdot}10^{-6}$	8.892	1 340 118	189 533



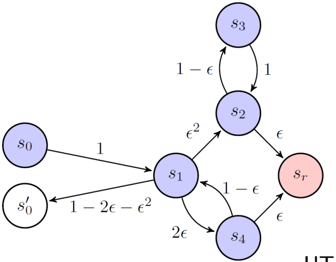
Case Study, $\epsilon = 0.1$, h = 95%-C.I. half-width

method	$\hat{p}_{\mathbb{Q}}$	$-\log(h)$	N	M
MC	$3.655 \cdot 10^{-2} \pm 1.21 \cdot 10^{-4}$	5.709	9 216 351	8 703 936
BFB	$3.654 \cdot 10^{-2} \pm 6.12 \cdot 10^{-5}$	6.392	3 600 591	2 223 172
BFB+	$3.66 \cdot 10^{-2} \pm 3.58 \cdot 10^{-5}$	6.93	3 659 717	2 287 933
BFB++	$3.657 \cdot 10^{-2} \pm 2.78 \cdot 10^{-5}$	6.946	3 678 719	2 297 880
ZVA	$3.659 \cdot 10^{-2} \pm 8.02 \cdot 10^{-6}$	8.425	2 281 743	323 533
ZVA+	$3.66 \cdot 10^{-2} \pm 2.55 \cdot 10^{-5}$	7.268	1 603 733	227 632
ZVA++	$3.659 \cdot 10^{-2} \pm 1.89 \cdot 10^{-6}$	8.892	1 340 118	189 533



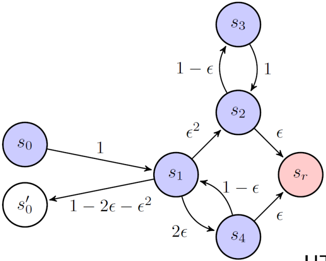
Case Study, $\epsilon = 0.001$, h = 95%-C.I. half-width

method	$\hat{p}_{\mathbb{Q}}$	$-\log(h)$	N	M
MC	$3.101 \cdot 10^{-6} \pm 9.86 \cdot 10^{-7}$	1.146	12 252 412	12 252 324
BFB	$2.966 \cdot 10^{-6} \pm 4.27 \cdot 10^{-8}$	4.241	56 153	34 867
BFB+	$3.006 \cdot 10^{-6} \pm 2.73 \cdot 10^{-10}$	9.307	59 460	37 166
BFB++	$3.006 \cdot 10^{-6} \pm 2.09 \cdot 10^{-10}$	9.336	59 245	36 888
ZVA	$3.006 \cdot 10^{-6} \pm 3.72 \cdot 10^{-11}$	11.3	50 430	81
ZVA+	$3.006 \cdot 10^{-6} \pm 1.2 \cdot 10^{-9}$	7.828	50 233	104
ZVA++	$3.006 \cdot 10^{-6} \pm 8.72 \cdot 10^{-14}$	14.29	38 439	84



Case Study, $\epsilon = 0.0001$, $h = 95\%$-C.I. half-width

method	$\hat{p}_{\mathbb{Q}}$	$-\log(h)$	N	M
MC	$0.0 \pm$ —	—	11 769 524	11 769 524
BFB	$3.021 \cdot 10^{-8} \pm 1.34 \cdot 10^{-9}$	3.115	5 840	3 595
BFB+	$3.001 \cdot 10^{-8} \pm 1.06 \cdot 10^{-12}$	10.25	3 404	2 134
BFB++	$3.001 \cdot 10^{-8} \pm 6.67 \cdot 10^{-13}$	10.49	6 044	3 833
ZVA	$3.001 \cdot 10^{-8} \pm 1.37 \cdot 10^{-13}$	12.29	3 653	1
ZVA+	$3.0 \cdot 10^{-8} \pm$ —	—	2 841	0
ZVA++	$3.0 \cdot 10^{-8} \pm$ —	—	2 696	0



Conclusions

Introduced new algorithm that can handle high-probability cycles. Gives vanishing relative error, improvement over Balanced Failure Biasing (BFB).

Extra variance reduction for BFB, also for Zero Variance Approximation if enough non-dominant paths are sampled.

Thank you for your attention.