Asymptotic Robustness of Estimators in Rare-Event Simulation

P. L'Ecuyer, Université de Montréal, Canada
J. H. Blanchet, Columbia University, USA
P. W. Glynn, Stanford University, USA
B. Tuffin, IRISA, Rennes, France

RESIM 2008, Rennes

Outline

- Rare-event setting and motivation.
- Asymptotic robustness properties: Definitions.
- Markov chain model and zero-variance approximation.
- Examples: Highly-reliable Markovian systems.
- Examples: Random walks.

Rare-event setting

We want to estimate a small quantity $\gamma = \gamma(\varepsilon) > 0$, where $\gamma(\varepsilon) \to 0$ when $\varepsilon \to 0$.

Rare-event setting

We want to estimate a small quantity $\gamma = \gamma(\varepsilon) > 0$, where $\gamma(\varepsilon) \to 0$ when $\varepsilon \to 0$.

We have an unbiased estimator (r.v.) $Y = Y(\varepsilon) \geq 0$.

Rare-event setting

We want to estimate a small quantity $\gamma = \gamma(\varepsilon) > 0$, where $\gamma(\varepsilon) \to 0$ when $\varepsilon \to 0$.

We have an unbiased estimator (r.v.) $Y = Y(\varepsilon) \geq 0$.

Example: $Y(\varepsilon)$ is an indicator function, $\mathbb{P}[Y(\varepsilon) = 1] = \gamma(\varepsilon)$.
Then the relative variance (squared relative error) blows up:

$$\frac{\mathrm{Var}[Y]}{\gamma^2(\varepsilon)} = \frac{1 - \gamma(\varepsilon)}{\gamma(\varepsilon)} \approx \frac{1}{\gamma(\varepsilon)} \to \infty \text{ when } \varepsilon \to 0.$$

Standard Monte Carlo (MC): estimate γ by $\bar{Y}_n$, the average of n i.i.d. copies of Y. For a meaningful estimate, need $n = \underline{O}(1/\gamma(\varepsilon))$.
If $\gamma = 10^{-10}$, for example, need $n = 10^{14}$ for 1% relative error.

Two popular cases, $\gamma(\varepsilon) \approx e^{-\eta/\varepsilon}$ and $\gamma(\varepsilon) \approx \mathrm{poly}(\varepsilon)$.

Some applications where this type of problem happens

- Expected amount of radiation that crosses a given protection shield.
- Probability of a large loss from an investment portfolio.
- Value-at-risk (quantile estimation).
- Ruin probability for an insurance firm.
- Probability that the completion time of a large project exceeds a given threshold.
- Probability of buffer overflow, or mean time to overflow, in a queueing system.
- Proportion of packets lost in a communication system.
- Air traffic control.
- Mean time to failure or other reliability or availability measure for a highly reliable system (e.g., fault-tolerant computers, safety systems).

Major techniques to improve this type of situation:
importance sampling (IS) and splitting.

Major techniques to improve this type of situation:
importance sampling (IS) and splitting.

Commonly-used robustness characterizations of $Y(\varepsilon)$:
It has bounded relative error (BRE) (bounded relative variance) if

$$\lim_{\varepsilon \to 0} \frac{\mathrm{Var}[Y(\varepsilon)]}{\gamma^2(\varepsilon)} < \infty.$$

It is logarithmically efficient (LE) or asymptotically optimal if

$$\lim_{\varepsilon \to 0} \frac{\ln \mathbb{E}[Y^2(\varepsilon)]}{2 \ln \gamma(\varepsilon)} = 1.$$

Means (roughly) that if $\gamma(\varepsilon) \to 0$ at an exponential rate, then the standard deviation converges at least at the same exponential rate.

Are there other useful characterizations?

To get a meaningful variance estimator when $\varepsilon \to 0$, we would like to control the relative error of the empirical variance. This involves the fourth moment of $Y(\varepsilon)$.

If we use a CLT, we may want to show that the quality of the normal approximation remains good when $\varepsilon \to 0$, by bounding the Berry-Esseen bound on the approx. error. This involves the third moment.

In some settings, we may be interested in bounding the relative moment of order $2 + \delta$ for some small $\delta > 0$.

And so on.

BRM-k

For $k \in [1, \infty)$, the relative moment of order k for $Y(\varepsilon)$ is

$$m_k(\varepsilon) = \mathbb{E}[Y^k(\varepsilon)]/\gamma^k(\varepsilon).$$

For any fixed ε, $m_k(\varepsilon)$ is nondecreasing in k (from Jensen).

BRM-k

For $k \in [1, \infty)$, the relative moment of order k for $Y(\varepsilon)$ is

$$m_k(\varepsilon) = \mathbb{E}[Y^k(\varepsilon)]/\gamma^k(\varepsilon).$$

For any fixed ε, $m_k(\varepsilon)$ is nondecreasing in k (from Jensen).

$Y(\varepsilon)$ has a bounded relative moment of order k (BRM-k) if

$$\limsup_{\varepsilon \to 0} m_k(\varepsilon) < \infty.$$

BRM-k

For $k \in [1, \infty)$, the relative moment of order k for $Y(\varepsilon)$ is

$$m_k(\varepsilon) = \mathbb{E}[Y^k(\varepsilon)]/\gamma^k(\varepsilon).$$

For any fixed ε, $m_k(\varepsilon)$ is nondecreasing in k (from Jensen).

$Y(\varepsilon)$ has a bounded relative moment of order k (BRM-k) if

$$\limsup_{\varepsilon \to 0} m_k(\varepsilon) < \infty.$$

Some properties:
(i) BRE is equivalent to BRM-2.
(ii) BRM-k implies BRM-k' for $1 \leq k' < k$.
(iii) For positive real numbers k, ℓ, m, if $Y(\varepsilon) = X^{\ell}(\varepsilon)$ is BRM-mk, then $Y'(\varepsilon) = X^{m\ell}(\varepsilon)$ is BRM-k.

LE-k

$Y(\varepsilon)$ has logarithmic efficiency of order k (LE-k) if

$$\lim_{\varepsilon \to 0} \frac{\ln \mathbb{E}[Y^k(\varepsilon)]}{k \ln \gamma(\varepsilon)} = 1.$$

Example:

$$\begin{aligned}\gamma(\varepsilon) &= \exp[-\eta/\varepsilon], \\ \mathbb{E}[Y^k(\varepsilon)] &= q(1/\varepsilon)\exp[-k\eta/\varepsilon]\end{aligned}$$

for some constant $\eta > 0$ and a slowly increasing function q (e.g., a polynomial). Then, we have LE-k but not BRE-k.

Example:

$$\begin{aligned}\gamma(\varepsilon) &= \exp[-\eta/\varepsilon], \\ \mathbb{E}[Y^k(\varepsilon)] &= q(1/\varepsilon)\exp[-k\eta/\varepsilon]\end{aligned}$$

for some constant $\eta > 0$ and a slowly increasing function q (e.g., a polynomial). Then, we have LE-k but not BRE-k.

Example:

$$\begin{aligned}\gamma^k(\varepsilon) &= q_1(\varepsilon) = \varepsilon^{t_1} + o(\varepsilon^{t_1}), \\ \mathbb{E}[Y^k(\varepsilon)] &= q_2(\varepsilon) = \varepsilon^{t_2} + o(\varepsilon^{t_2}),\end{aligned}$$

where $t_2 \le t_1$. Here,

$$\lim_{\varepsilon\to 0} \frac{\ln \mathbb{E}[Y^k(\varepsilon)]}{k \ln \gamma(\varepsilon)} = \frac{t_2}{t_1}.$$

We have BRM-k iff $t_2 = t_1$ iff LE-k.

Bounded normal approximation

Berry-Esseen Theorem (one version):
$Y_1, \ldots, Y_n$ i.i.d. r.v.'s with $\mathbb{E}[Y_1] = 0$, $\mathrm{Var}[Y_1] = \sigma^2$, $\mathbb{E}[|Y_1|^3] = \beta^3$.
Let $\bar{Y}_n$ and S_n^2 be the empirical mean and variance, and F_n the distribution function of $\sqrt{n}\bar{Y}_n/S_n$. Then, there is an absolute constant $a < \infty$ such that

$$\sup_{n \geq 2,\, x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{a\beta_3}{\sigma^3\sqrt{n}}.$$

$Y(\varepsilon)$ is said to have Bounded Normal Approximation (BNA) if

$$\limsup_{\varepsilon \to 0} \frac{\mathbb{E}\left[|Y(\varepsilon) - \gamma(\varepsilon)|^3\right]}{\sigma^3(\varepsilon)} < \infty$$

(Tuffin 1999). This requires that the Berry-Esseen bound remains $O(n^{-1/2})$ when $\varepsilon \to 0$.

BNA is not equivalent to BRM-3, because we divide by $\sigma^3(\varepsilon)$ here. One can have BNA and not BRM-3, or vice-versa.

There are more general versions of the Berry-Esseen inequality that require only a bounded moment of order $2+\delta$ for any $\delta \in (0,1]$ instead of the third moment β_3; see, e.g., Petrov (1995). But the bound on $|F_n(x) - \Phi(x)|$ is only $O(n^{-\delta/2})$ instead of $O(n^{-1/2})$.

Robustness of the empirical variance

Let $X_1(\varepsilon), \ldots, X_n(\varepsilon)$ be an i.i.d. sample of $X(\varepsilon)$. We consider the empirical variance $Y(\varepsilon) = S_n^2(\varepsilon)$ as an estimator of the variance $\sigma^2(\varepsilon)$. Let $\gamma(\varepsilon) = \mathbb{E}[X(\varepsilon)]$.

Robustness of the empirical variance

Let $X_1(\varepsilon), \ldots, X_n(\varepsilon)$ be an i.i.d. sample of $X(\varepsilon)$. We consider the empirical variance $Y(\varepsilon) = S_n^2(\varepsilon)$ as an estimator of the variance $\sigma^2(\varepsilon)$. Let $\gamma(\varepsilon) = \mathbb{E}[X(\varepsilon)]$.

BRM-4 for $X(\varepsilon)$ and BRE for $S_n^2(\varepsilon)$ are both linked to $\mathbb{E}[X^4(\varepsilon)]$, but they are not equivalent. We have

$$\frac{\mathrm{Var}[S_n^2(\varepsilon)]}{\sigma^4(\varepsilon)} = \Theta\left(\frac{\mathbb{E}[(X(\varepsilon) - \gamma(\varepsilon))^4]}{\sigma^4(\varepsilon)}\right)$$

which differs in general from

$$\Theta\left(\mathbb{E}[X^4(\varepsilon)]/\gamma^4(\varepsilon)\right).$$

Proposition: If $\sigma^2(\varepsilon) = \Theta(\gamma^2(\varepsilon))$, then BRM-$2k$ for $X(\varepsilon)$ implies BRM-k for $S_n^2(\varepsilon)$, for any $k \geq 1$.

A similar observation applies to the equivalence between LE-4 for $X(\varepsilon)$ and LE for $S_n^2(\varepsilon)$.

Vanishing relative centered moments

The relative centered moment of order k, for $Y(\varepsilon)$, is

$$c_k(\varepsilon) = \frac{\mathbb{E}[|Y(\varepsilon) - \gamma(\varepsilon)|^k]}{\gamma^k(\varepsilon)}.$$

$Y(\varepsilon)$ has vanishing relative centered moment of order k (VRCM-k) if

$$\limsup_{\varepsilon \to 0} c_k(\varepsilon) = 0.$$

True if and only if

$$\limsup_{\varepsilon \to 0} m_k(\varepsilon) = 1.$$

Vanishing relative centered moments

The relative centered moment of order k, for $Y(\varepsilon)$, is

$$c_k(\varepsilon) = \frac{\mathbb{E}[|Y(\varepsilon) - \gamma(\varepsilon)|^k]}{\gamma^k(\varepsilon)}.$$

$Y(\varepsilon)$ has vanishing relative centered moment of order k (VRCM-k) if

$$\limsup_{\varepsilon \to 0} c_k(\varepsilon) = 0.$$

True if and only if

$$\limsup_{\varepsilon \to 0} m_k(\varepsilon) = 1.$$

It has vanishing relative variance or relative error (VRE), if

$$\limsup_{\varepsilon \to 0} \frac{\sigma(\varepsilon)}{\gamma(\varepsilon)} = 0.$$

Vanishing relative centered moments

The relative centered moment of order k, for $Y(\varepsilon)$, is

$$c_k(\varepsilon) = \frac{\mathbb{E}[|Y(\varepsilon) - \gamma(\varepsilon)|^k]}{\gamma^k(\varepsilon)}.$$

$Y(\varepsilon)$ has vanishing relative centered moment of order k (VRCM-k) if

$$\limsup_{\varepsilon \to 0} c_k(\varepsilon) = 0.$$

True if and only if

$$\limsup_{\varepsilon \to 0} m_k(\varepsilon) = 1.$$

It has vanishing relative variance or relative error (VRE), if

$$\limsup_{\varepsilon \to 0} \frac{\sigma(\varepsilon)}{\gamma(\varepsilon)} = 0.$$

VRCM-k implies VRCM-k' for $1 \leq k' \leq k$. It also implies BRM-k.

VRCM implies convergence to the zero-variance IS

Suppose

$$\gamma(\varepsilon) = \mathbb{E}_{P_\varepsilon}[Y(\varepsilon)] = \int_\Omega Y(\varepsilon, \omega) dP_\epsilon(\omega).$$

Here, we can get zero variance (in principle) by doing importance sampling with the measure Q_ε^* defined by

$$dQ_\varepsilon^*(\omega) = \frac{Y(\varepsilon, \omega)}{\gamma(\varepsilon)} dP_\varepsilon(\omega).$$

Proposition: If $Y(\varepsilon)$ is VRCM-$(1+\delta)$ for some $\delta > 0$, then

$$\lim_{\varepsilon \to 0} \sup_A |P_\varepsilon(A) - Q_\varepsilon^*(A)| = 0.$$

Proof: For any measurable set A,

$$\begin{aligned}
\sup_A |Q^*_\varepsilon(A) - P_\varepsilon(A)| &\leq \sup_A |\mathbb{E}_{P_\varepsilon}[(dQ^*_\varepsilon/dP_\varepsilon)\,\mathbb{I}(A)] - \mathbb{E}_{P_\varepsilon}[\mathbb{I}(A)]| \\
&\leq \mathbb{E}_{P_\varepsilon}|dQ^*_\varepsilon/dP_\varepsilon - 1| \\
&\leq \mathbb{E}^{1/(1+\delta)}_{P_\varepsilon}\left[|dQ^*_\varepsilon/dP_\varepsilon - 1|^{(1+\delta)}\right] \\
&\leq \mathbb{E}^{1/(1+\delta)}_{P_\varepsilon}\left[|Y(\varepsilon)/\gamma(\varepsilon) - 1|^{(1+\delta)}\right] \\
&= [c_{1+\delta}(\varepsilon)]^{1/(1+\delta)} \\
&\xrightarrow{\varepsilon\to 0} 0.
\end{aligned}$$

First-Passage Probability in a Markov Chain

Markov chain $X = \{X_j, j \geq 0\}$ with state space $\mathcal{S}$ and transition kernel $K = \{K(x, C) : x \in S, C \subseteq \mathcal{S}\}$.

For $C \subset \mathcal{S}$, define $\tau_C = \inf\{j \geq 0 : X_j \in C\}$.
Given A and B, two disjoint subsets of $\mathcal{S}$, and initial state $x_0 \in (A \cup B)^c$, want to estimate $\gamma(x_0)$, where

$$\gamma(x) = \gamma(x, \varepsilon) = \mathbb{P}[\tau_B < \tau_A \mid X_0 = x]$$

We have $\gamma(x) = 1$ for $x \in B$ and $\gamma(x) = 0$ for $x \in A$.
Here, K, A, and B may depend on ε.

Importance sampling (IS) replace K by another kernel.
We get zero variance if we use the kernel

$$K^*(x, dy) = K(x, dy)\,\frac{\gamma(y)}{\gamma(x)}.$$

Importance sampling (IS) replace K by another kernel.
We get zero variance if we use the kernel

$$K^*(x, dy) = K(x, dy)\,\frac{\gamma(y)}{\gamma(x)}.$$

Not practical. But can be used to design a good IS scheme of the form

$$K_v(x, dy) = K(x, dy)\,\frac{v(y)}{w(x)},$$

where $v : \mathcal{S} \to [0, \infty)$ is a good approximation of $\gamma(\cdot)$ and

$$w(x) = \int_{\mathcal{S}} K(x, dy)\, v(y)$$

is the appropriate normalizing constant, assumed finite for $x \in (A \cup B)^c$. (When $v = \gamma$, we have $w = v$.)

IS estimator of $\gamma(x_0)$:

$$Y = Y(\varepsilon) = \mathbb{I}[\tau_B < \tau_A]\, L,$$

where

$$L = \prod_{k=1}^{\tau_B} \frac{w(X_{k-1})}{v(X_k)} = \frac{w(X_0)}{v(X_{\tau_B})} \prod_{k=1}^{\tau_B - 1} \frac{w(X_k)}{v(X_k)}$$

Can take $v(x) = 1$ for $x \in B$ and $v(x) = 0$ for $x \in A$.

IS estimator of $\gamma(x_0)$:

$$Y = Y(\varepsilon) = \mathbb{I}[\tau_B < \tau_A] \, L,$$

where

$$L = \prod_{k=1}^{\tau_B} \frac{w(X_{k-1})}{v(X_k)} = \frac{w(X_0)}{v(X_{\tau_B})} \prod_{k=1}^{\tau_B - 1} \frac{w(X_k)}{v(X_k)}$$

Can take $v(x) = 1$ for $x \in B$ and $v(x) = 0$ for $x \in A$.

To establish robustness properties such as LE-k, BRM-k, and VRCM-k, we need an asymptotic bound on $\mathbb{E}[Y^k(\varepsilon)]/\gamma^k(x_0, \varepsilon)$.

Proposition: Bounds via Lyapunov inequalities.

Suppose there are positive constants κ_1 and κ_2 and a function $h_k : \mathcal{S} \to [0, \infty)$ such that $v(x) \geq \kappa_1$ and $h_k(x) \geq \kappa_2$ for each $x \in B$, and

$$\mathbb{E}_{v,x}\left[\left(\frac{w(x)}{v(x)}\right)^k h_k(X_1)\right] \leq h_k(x)$$

for all $x \in (A \cup B)^c$.

Then, for all $x \in (A \cup B)^c$,

$$\mathbb{E}_{v,x}[Y^k] \leq \frac{v^k(x) h_k(x)}{\kappa_1 \kappa_2}.$$

Corollary.

Under the proposition's conditions:

(i) If
$$\lim_{\varepsilon \to 0} k \ln[v(x_0, \varepsilon)/\gamma(x_0, \varepsilon)] + h_k(x_0, \varepsilon) = 1,$$
then $Y(\varepsilon)$ is LE-k.

(ii) If
$$\lim_{\varepsilon \to 0} [v(x_0, \varepsilon)/\gamma(x_0, \varepsilon)]^k h_k(x_0, \varepsilon) < \infty,$$
then $Y(\varepsilon)$ is BRM-k.

(iii) If
$$\lim_{\varepsilon \to 0} \frac{[v(x_0, \varepsilon)/\gamma(x_0, \varepsilon)]^k h_k(x_0, \varepsilon)}{\kappa_1 \kappa_2} = 1,$$
then $Y(\varepsilon)$ is VRCM-k.

HRMS-type Framework

Suppose $\mathcal{S}$ is finite and $\{X_j, j \geq 0\}$ has transition probabilities

$$p(x, y, \varepsilon) = \mathbb{P}[X_j = y \mid X_{j-1} = x] = a(x, y)\varepsilon^{b(x,y)},$$

where $a(x, y)$ and $b(x, y)$ are non-negative constants.

Ex.: multicomponent highly-reliable Markovian system (HRMS):
$A = \{x_0\}$ is the single state where all components are up and B is the set of states where the system is failed.
We typically have $b(x, y) > 0$ for "failure" transitions and $b(x, y) = 0$ for "repair" transitions (Shahabuddin, Nakayama, ...).
In this setting, $\gamma(\varepsilon) = \Theta(\varepsilon^r)$ for some $r > 0$.

Some proposed IS heuristics:

Balanced failure biasing (BFB) (Shahabuddin 1994) changes p to q as follows, for $x \notin B$:

$$q(x,y) = \begin{cases} \frac{1}{|F(x)|} & \text{if } y \in F(x) \text{ and } p_{\mathrm{R}}(x) = 0; \\ \rho \, \frac{1}{|F(x)|} & \text{if } y \in F(x) \text{ and } p_{\mathrm{R}}(x) > 0; \\ (1-\rho)\, \frac{p(x,y)}{p_{\mathrm{R}}(x)} & \text{if } y \in R(x); \\ 0 & \text{otherwise.} \end{cases}$$

Some proposed IS heuristics:

Balanced failure biasing (BFB) (Shahabuddin 1994) changes p to q as follows, for $x \notin B$:

$$q(x,y) = \begin{cases} \frac{1}{|F(x)|} & \text{if } y \in F(x) \text{ and } p_{\mathrm{R}}(x) = 0; \\ \rho\ \frac{1}{|F(x)|} & \text{if } y \in F(x) \text{ and } p_{\mathrm{R}}(x) > 0; \\ (1-\rho)\ \frac{p(x,y)}{p_{\mathrm{R}}(x)} & \text{if } y \in R(x); \\ 0 & \text{otherwise.} \end{cases}$$

Simple failure biasing (SFB) (Shahabuddin 1988): Replace $1/|F(x)|$ above by $p(x,y)/\sum_{y \in F(x)} p(x,y)$.

Some proposed IS heuristics:

Balanced failure biasing (BFB) (Shahabuddin 1994) changes p to q as follows, for $x \notin B$:

$$q(x,y) = \begin{cases} \frac{1}{|F(x)|} & \text{if } y \in F(x) \text{ and } p_{\mathrm{R}}(x) = 0; \\ \rho \, \frac{1}{|F(x)|} & \text{if } y \in F(x) \text{ and } p_{\mathrm{R}}(x) > 0; \\ (1-\rho) \, \frac{p(x,y)}{p_{\mathrm{R}}(x)} & \text{if } y \in R(x); \\ 0 & \text{otherwise.} \end{cases}$$

Simple failure biasing (SFB) (Shahabuddin 1988): Replace $1/|F(x)|$ above by $p(x,y)/\sum_{y \in F(x)} p(x,y)$.

SBLR (Alexopoulos and Shultes 2001) changes the probabilities in a way that over any cycle in the visited states during the simulation, the cumulated likelihood ratio remains bounded

Some proposed IS heuristics:

Balanced failure biasing (BFB) (Shahabuddin 1994) changes p to q as follows, for $x \notin B$:

$$q(x,y) = \begin{cases} \frac{1}{|F(x)|} & \text{if } y \in F(x) \text{ and } p_{\mathrm{R}}(x) = 0; \\ \rho \, \frac{1}{|F(x)|} & \text{if } y \in F(x) \text{ and } p_{\mathrm{R}}(x) > 0; \\ (1-\rho) \, \frac{p(x,y)}{p_{\mathrm{R}}(x)} & \text{if } y \in R(x); \\ 0 & \text{otherwise.} \end{cases}$$

Simple failure biasing (SFB) (Shahabuddin 1988): Replace $1/|F(x)|$ above by $p(x,y)/\sum_{y \in F(x)} p(x,y)$.

SBLR (Alexopoulos and Shultes 2001) changes the probabilities in a way that over any cycle in the visited states during the simulation, the cumulated likelihood ratio remains bounded

These methods do not attempt to mimic zero-variance sampling.

Proposed approximation (ZVA)

Approximate γ by some easily computable function v, and plug into zero-variance formula.

For any state $y \notin B$, let $\Gamma(y)$ be the set of all paths $\pi = (y = y_0 \to y_1 \to \cdots \to y_k)$ where $y_1, \ldots, y_{k-1} \notin B \cup \{\mathbf{0}\}$, $y_k \in B$, and having positive probability

$$p(\pi) = \prod_{j=1}^{k} p(y_{j-1}, y_j) > 0.$$

Proposed approximation (ZVA)

Approximate γ by some easily computable function v, and plug into zero-variance formula.

For any state $y \notin B$, let $\Gamma(y)$ be the set of all paths $\pi = (y = y_0 \to y_1 \to \cdots \to y_k)$ where $y_1, \ldots, y_{k-1} \notin B \cup \{\mathbf{0}\}$, $y_k \in B$, and having positive probability

$$p(\pi) = \prod_{j=1}^{k} p(y_{j-1}, y_j) > 0.$$

Because these paths represent disjoint events, we have

$$\gamma(y) = \sum_{\pi \in \Gamma(y)} p(\pi).$$

Proposed approximation (ZVA)

Approximate γ by some easily computable function v, and plug into zero-variance formula.

For any state $y \notin B$, let $\Gamma(y)$ be the set of all paths $\pi = (y = y_0 \to y_1 \to \cdots \to y_k)$ where $y_1, \ldots, y_{k-1} \notin B \cup \{\mathbf{0}\}$, $y_k \in B$, and having positive probability

$$p(\pi) = \prod_{j=1}^{k} p(y_{j-1}, y_j) > 0.$$

Because these paths represent disjoint events, we have

$$\gamma(y) = \sum_{\pi \in \Gamma(y)} p(\pi).$$

This last sum may contain a huge (perhaps ∞) number of terms.

A very crude approximation is to just take the path with largest probability, i.e., approximate

$$\gamma(y) = \sum_{\pi \in \Gamma(y)} p(\pi)$$

by its lower bound

$$v_0(y) = \max_{\pi \in \Gamma(y)} p(\pi).$$

Computing $v_0(y)$ amounts to computing a shortest path from y to B, where the length of a link $y' \to y''$ is $-\log p(y', y'')$. Easy.

A very crude approximation is to just take the path with largest probability, i.e., approximate

$$\gamma(y) = \sum_{\pi \in \Gamma(y)} p(\pi)$$

by its lower bound

$$v_0(y) = \max_{\pi \in \Gamma(y)} p(\pi).$$

Computing $v_0(y)$ amounts to computing a shortest path from y to B, where the length of a link $y' \to y''$ is $-\log p(y', y'')$. Easy.

This would work fine if a single path dominates the sum (this may happen when failure transitions have very small probabilities), but this v_0 may underestimate the bound significantly.

Slight improvements: take the sum of probabilities of a few disjoint paths.

Refinements

Typically, the farther we are from B, the more v_0 underestimates γ. Close to B, things are fine, but not close to $\mathbf{0}$.

First simple correction:

1. Estimate $\gamma(\mathbf{0})$ in preliminary runs with crude IS strategy;
2. Find constant $\alpha \leq 1$ such that $(v_0(\mathbf{0}))^\alpha$ equals this estimate;
3. Use $v_1(y) = (v_0(y))^\alpha$ for all $y \notin B$ as approx. of $\gamma(y)$.

This v_1 matches γ for $y \in B$ and matches its estimate at $y = \mathbf{0}$.

Refinements

Typically, the farther we are from B, the more v_0 underestimates γ. Close to B, things are fine, but not close to $\mathbf{0}$.

First simple correction:

1. Estimate $\gamma(\mathbf{0})$ in preliminary runs with crude IS strategy;
2. Find constant $\alpha \leq 1$ such that $(v_0(\mathbf{0}))^{\alpha}$ equals this estimate;
3. Use $v_1(y) = (v_0(y))^{\alpha}$ for all $y \notin B$ as approx. of $\gamma(y)$.

This v_1 matches γ for $y \in B$ and matches its estimate at $y = \mathbf{0}$.

Second refinement: Replace α by a state-dependent exponent

$$\alpha(y) = 1 + [\alpha(\mathbf{0}) - 1] \frac{\log v_0(y)}{\log v_0(\mathbf{0})},$$

where $\alpha(\mathbf{0}) = \alpha$ as above. This $\alpha(y)$ changes progressively from 1 near B to $\alpha(\mathbf{0}) < 1$ in state $\mathbf{0}$. The correction here is milder than in the previous case when we are close to B.
Let $v_2(y) = (v_0(y))^{\alpha(y)}$ be the resulting approximation.

Example: Three types of components

$c = 3$ and $n_1 = n_2 = n_3$.
Expon. repair times with mean 1.
Failure rate λ_i for component type i,
with $\lambda_1 = \varepsilon$, $\lambda_2 = 1.5\varepsilon$, and $\lambda_3 = 2\varepsilon^2$, for some small real number ε.

We will try different values of (n_i, ε).

B = states where at least one component type has fewer than 2 operational units.

Example: Three types of components

$c = 3$ and $n_1 = n_2 = n_3$.
Expon. repair times with mean 1.
Failure rate λ_i for component type i,
with $\lambda_1 = \varepsilon$, $\lambda_2 = 1.5\varepsilon$, and $\lambda_3 = 2\varepsilon^2$, for some small real number ε.

We will try different values of (n_i, ε).

B = states where at least one component type has fewer than 2 operational units.

To define $v_0(y)$, we consider all three paths to B that result from failures of a single component type, and sum their probabilities.

The table contains results with $n = 2^{20}$ runs.

Best estimate of $\gamma(\mathbf{0})$: obtained from a large number of runs with our best IS strategies.

Mean

n_i	ε	$\gamma(\mathbf{0})$	$v_0(\mathbf{0})$	BFB	SBLR
3	0.001	2.6×10^{-3}	1.3×10^{-3}	2.7×10^{-3}	2.6×10^{-3}
6	0.01	1.8×10^{-7}	3.4×10^{-8}	1.9×10^{-7}	$[9.9 \times 10^{-7}]$
6	0.001	1.7×10^{-11}	3.4×10^{-12}	1.8×10^{-11}	(1.8×10^{-16})
12	0.1	6.0×10^{-8}	3.2×10^{-9}	4.8×10^{-8}	1.3×10^{-8}
12	0.001	3.9×10^{-28}	3.5×10^{-29}	(1.8×10^{-40})	(2.9×10^{-45})

Variance

n_i	ε	BFB	SBLR
3	0.001	1.8×10^{-2}	8.0×10^{-3}
6	0.01	6.3×10^{-11}	(4.5×10^{-16})
6	0.001	8.8×10^{-19}	(2.0×10^{-26})
12	0.1	8.1×10^{-10}	1.7×10^{-10}
12	0.001	(3.2×10^{-74})	(3.5×10^{-84})

Mean

n_i	ε	$\gamma(\mathbf{0})$	ZVA(v_0)	ZVA(v_1)	ZVA(v_2)
3	0.001	2.6×10^{-3}	2.6×10^{-3}	2.6×10^{-3}	2.6×10^{-3}
6	0.01	1.8×10^{-7}	1.8×10^{-7}	1.8×10^{-7}	1.8×10^{-7}
6	0.001	1.7×10^{-11}	1.7×10^{-11}	1.7×10^{-11}	1.7×10^{-11}
12	0.1	6.0×10^{-8}	6.0×10^{-8}	6.2×10^{-8}	6.7×10^{-8}
12	0.001	3.9×10^{-28}	3.9×10^{-28}	3.9×10^{-28}	3.9×10^{-28}

Variance

n_i	ε	α	ZVA(v_0)	ZVA(v_1)	ZVA(v_2)	RE(v_2)
3	0.001	0.906	6.5×10^{-4}	2.7×10^{-3}	9.3×10^{-9}	0.04
6	0.01	0.903	2.0×10^{-14}	1.2×10^{-14}	7.7×10^{-15}	0.48
6	0.001	0.939	1.2×10^{-23}	1.1×10^{-23}	7.6×10^{-24}	0.16
12	0.1	0.851	1.6×10^{-10}	2.9×10^{-10}	1.5×10^{-11}	64.50
12	0.001	0.963	1.4×10^{-55}	9.3×10^{-56}	9.4×10^{-56}	0.78

Proposition. In this HRMS framework, with SFB, BRM-k and LE-k are equivalent.
They are also equivalent for the gth empirical moment.

Proposition. In this HRMS framework, with SFB, BRM-k and LE-k are equivalent.
They are also equivalent for the gth empirical moment.

Proposition. Suppose we have an IS scheme such that $p(x, y, \varepsilon) = \Theta(\varepsilon^d)$ implies $q(x, y, \varepsilon) = \Theta(\varepsilon^\ell)$ for $\ell \leq d$. Let $\mathbb{E}[Y^g(\varepsilon)] = \Theta(\varepsilon^{s_g})]$. Then we have BRM-$k$ of the g-th empirical moment if and only if for all integers m such that $r \leq m < ks_g$ and all sample paths $(x_0, \cdots, x_n)$ leading to B and having probability $\Theta(\varepsilon^m)$,

$$\mathbb{P}^*\{(X_0, \cdots, X_\tau) = (x_0, \cdots, x_n)\} = \Theta(\varepsilon^\ell)$$

for some $\ell \leq k(mg - s_g)/(kg - 1)$.

Proposition. In this HRMS framework, with SFB, BRM-k and LE-k are equivalent.
They are also equivalent for the gth empirical moment.

Proposition. Suppose we have an IS scheme such that $p(x, y, \varepsilon) = \Theta(\varepsilon^d)$ implies $q(x, y, \varepsilon) = \Theta(\varepsilon^\ell)$ for $\ell \leq d$. Let $\mathbb{E}[Y^g(\varepsilon)] = \Theta(\varepsilon^{s_g})]$. Then we have BRM-$k$ of the g-th empirical moment if and only if for all integers m such that $r \leq m < ks_g$ and all sample paths $(x_0, \cdots, x_n)$ leading to B and having probability $\Theta(\varepsilon^m)$,

$$\mathbb{P}^*\{(X_0, \cdots, X_\tau) = (x_0, \cdots, x_n)\} = \Theta(\varepsilon^\ell)$$

for some $\ell \leq k(mg - s_g)/(kg - 1)$.

Proposition. With SFB or BFB, one cannot achieve VRCM-k.

ZVA

Proposition. With our ZVA scheme, if we just take $v(y)$ as the probability of the most probable path to failure from y, then $v(y) = \Theta(\gamma(y))$ for all y and we have BRE.

ZVA

Proposition. With our ZVA scheme, if we just take $v(y)$ as the probability of the most probable path to failure from y, then $v(y) = \Theta(\gamma(y))$ for all y and we have BRE.

Proposition. If $v(y)/\gamma(y) \to 1$ for each y when $\varepsilon \to 0$, then we have VRCM-k. This holds if $v(y)$ is the sum of probabilities of all the dominant paths from y (those with the smallest power of ε).

Probability of a Large Deviation for a Random Walk

Let $D_1, D_2, \ldots$ be i.i.d. random variables, and $S_j = D_1 + \cdots + D_j$ for $j \geq 0$. Let $\ell > \mathbb{E}[D_j]$, $n = n(\varepsilon) = \lceil 1/\varepsilon \rceil$, and

$$\gamma(\varepsilon) = \mathbb{P}[S_n/n \geq \ell].$$

We have $\gamma(\varepsilon) \to 0$ when $\varepsilon \to 0$.

State-Independent Exponential Twisting

Here, it is well known that an LE-2 estimator can be obtained via IS with exponential twisting, if D_j has a light tail distribution

Idea: multiply the density of D_j by $e^{\theta x}$ and normalize.
The IS estimator is $Y(\theta, \varepsilon) = \mathbb{I}[S_n \geq n\ell] \cdot L(\theta, S_n)$.

State-Independent Exponential Twisting

Here, it is well known that an LE-2 estimator can be obtained via IS with exponential twisting, if D_j has a light tail distribution

Idea: multiply the density of D_j by $e^{\theta x}$ and normalize.
The IS estimator is $Y(\theta, \varepsilon) = \mathbb{I}[S_n \geq n\ell] \cdot L(\theta, S_n)$.

Let θ_ℓ^* be a positive root of $\Psi'(\theta) = \ell$, where Ψ is the cumulant generating function of D_j.
Let $I(\ell) = \ell\theta_\ell^* - \Psi(\theta_\ell^*)$, the large deviation rate function.

Results of Sadowsky (1993) directly imply the following:

Proposition. For any integer $k \geq 2$ and any θ, $Y(\theta, \varepsilon)$ is not BRM-k. It is LE-k if and only if $\theta = \theta_\ell^*$.

Proposition. The sample variance of m copies of $Y(\theta, \varepsilon)$, as an estimator of the true variance, is not BRM-k.
It is LE-k if and only if $\theta = \theta_\ell^*$.

A State-Dependent IS Scheme Can Achieve BRM-k

We know that for $n \to \infty$,

$$\gamma(\varepsilon) = \mathbb{P}[S_n \geq n\ell] = \frac{\exp[-nI(\ell)]}{[2\pi n \Psi''(\theta_\ell^*)]^{1/2}\theta_\ell^*}[1 + o(1)].$$

A State-Dependent IS Scheme Can Achieve BRM-k

We know that for $n \to \infty$,

$$\gamma(\varepsilon) = \mathbb{P}[S_n \geq n\ell] = \frac{\exp[-nI(\ell)]}{[2\pi n \Psi''(\theta_\ell^*)]^{1/2}\theta_\ell^*}[1 + o(1)].$$

When the Markov chain is in state $X_j = (j, S_j) = (j, s) = x$, we can approximate $\gamma(x) = \gamma(j, s) = \mathbb{P}[S_n - S_j \geq \ell - s]$ by

$$v(j, s) = \frac{\exp[-(n-j)I(\ell_j)]}{[2\pi(n-j)\Psi''(\theta_{\ell_j}^*)]^{1/2}\theta_{\ell_j}^*}$$

where $\ell_j = (n\ell - S_j)/(n - j)$.
Then we use the corresponding zero-variance approximation.

By verifying the Lyapunov conditions, we can show that this IS estimator has BRM-k for all k.

Increments With Heavy Tails

Random walk with negative drift; probability that the walk exceeds a given positive threshold.

Can obtain state-dependent IS having BRM-k for all k.