An Analysis of the LRE-Algorithm using Sojourn Times

Norbert Th. Müller
Abteilung Informatik — Universität Trier
D-54286 Trier, Germany
E-mail: mueller@uni-trier.de
Tel: ++49-651-201-2845
Fax: ++49-651-201-3805

KEYWORDS

Discrete event simulation, LRE-algorithm, confidence interval, sojourn times

ABSTRACT

The classical method of evaluation of simulations surely is the batch means method, see e.g. [Bratley et al. 1987], giving confidence intervals to express the precision of the simulation results. Unfortunately, it is not suited for rare event simulation techniques, as they do not produce the necessary large batches.
As an alternative, the LRE-algorithm has been introduced in [Schreiber 1984], where the lengthy final analysis of its basis has just been given in [Schreiber 1999]. Its name is based on the 'Limited Relative Error' that has been chosen to describe the precision of the results.
In this paper, we present another, much simpler, analysis of the LRE-algorithm based on sojourn times, which gives a better understanding of the algorithm and clearly exhibits certain limitations. Additionally, it allows to create confidence intervals, and so it is possible to compare the results of the two very different methods.

The LRE-algorithm

The batch means method (e.g. [Bratley et al. 1987]) tries to construct independent samples from a given time series by building batches that tend to have lesser autocorrelation than the original series. So essentially it aims on the reduction of autocorrelation. It results in confidence intervals for the mean of the simulation results.
In contrast, the LRE-algorithm, as introduced in e.g. in [Schreiber 1984] and used in a variety of papers later on, e.g. [Schreiber 1999, Görg and Schreiber 1996], measures the autocorrelation and tries to deduce the precision of the simulation results from this autocorrelation (instead of reducing it).
In addition, it does not aim at producing mean values, but tries to estimate the distribution of the results. Additionally, instead of confidence intervals, the 'Limited Relative Error' has been introduced.
In the following, we will briefly recall the LRE-algorithm, more precisely: version LRE-III for discrete sequences of real valued random variables $X_1, X_2, X_3, \ldots$. We assume that these values are identically distributed with $F(x) := P(X_i \leq x)$ independent from the index i, but of course there may be significant autocorrelation.
For simplicity, we just concentrate on the estimation of one point of the distribution of the X_i, as this also is the starting point of the original analysis. So for one chosen value x, we try to estimate $F(x)$ or, equivalently, the inverse distribution function $G(x) = 1 - F(x)$.
The LRE-algorithm essentially reduces the original time series $X_1, X_2, X_3, \ldots$ to a sequence $Y_1, Y_2, Y_3, \ldots$ of boolean values with $Y_i = 1 \Leftrightarrow X_i \leq x$. Please note that the random variables Y_i of this new time series are still identically distributed and still may be autocorrelated! For the expectation $E[Y]$ of the Y_i, we get

$$\begin{aligned} E[Y] &= 1 \cdot P(X \leq x) + 0 \cdot P(X > x) \\ &= P(X \leq x) = F(x) \end{aligned}$$

So our original question of estimating $F(x)$ has been reduced to the estimation of $E[Y]$. To do this, the following values are computed from the time series (Y_n) (i.e. they are measured during a simulation):

- quantity $r = r(n)$ of values $Y_i = 1$ for $1 \leq i \leq n$, i.e. the quantity of X_i with $X_i \leq x$.
- quantity $a = a(n)$ of transitions from $Y_i = 1$ to $Y_{i+1} = 0$ for $1 \leq i \leq n$, i.e. the quantity of observed pairs (X_i, X_{i+1}) with $X_i \leq x < X_{i+1}$.

In this chapter, n will be treated as fixed, and we will simply use a and r instead of $a(n)$ and $r(n)$.
Between any two subsequent transitions $Y_i = 1 \to Y_{i+1} = 0$ and $Y_j = 1 \to Y_{j+1} = 0$ there must be a third transition $Y_k = 0 \to Y_{k+1} = 1$, $i < k < j$. So it is not necessary to measure the following values, as they can be deduced with sufficient precision from n, a, r:

- quantity $v = n - r$ of values $Y_i = 0$ for $1 \leq i \leq n$, i.e. the quantity of X_i with $X_i > x$.
- quantity $c \approx a$ of transitions from $Y_i = 0$ to $Y_{i+1} = 1$ for $1 \leq i \leq n$, i.e. the quantity of observed pairs (X_i, X_{i+1}) with $X_i > x \geq X_{i+1}$.
- quantity $b \approx r - a$ of transitions within $\{X \leq x\}$, i.e. the quantity of pairs (Y_i, Y_{i+1}) with $Y_i = Y_{i+1} = 1$.

- quantity $d \approx v - a$ of transitions within $\{X > x\}$, i.e. the quantity of pairs (Y_i, Y_{i+1}) with $Y_i = Y_{i+1} = 0$.

Please note that the approximations from above may all be wrong by at most 1. So in the following we will simply use $c = a$, $b = r - a$ and $d = v - a = n - r - a$.
The following approximation is immediate:

$$F(x) = E[Y] \approx r/n \tag{1}$$

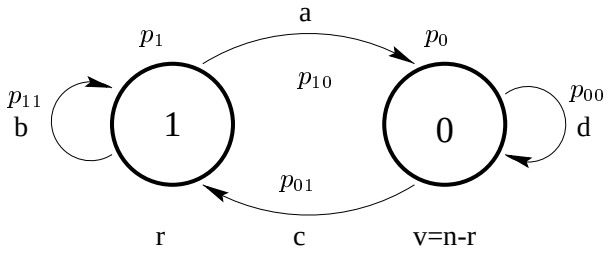
To estimate the precision of this approximation, the sequence (Y_i) is treated like a two node chain. Then the central assumption of the LRE-algorithm is as follows:

- (1-LRE) Assume the nodes 0 and 1 to be memoryless, i.e. treat the system like a discrete homogenous Markov chain with just the two nodes 0 and 1!

In [Görg and Schreiber 1996, Schreiber 1999] this chain is called the $F(x)$-equivalent 2-Node Markov chain. This chain is characterized by its one step transition probabilities:

$$p_{ij} = \text{Prob(transitions starting in } i \text{ lead to } j)$$

where $i, j \in \{0, 1\}$. Resulting from this, there are the steady state probabilities p_0 and p_1 for being in state 0 resp. in state 1. So the central assumption of the LRE algorithm is $F(x) = p_1$ and $1 - F(x) = p_0$ for these values derived from the Markov chain.
Later in this paper we will consider the implications of this central assumption and also discuss its validity. But first we recall the results from [Schreiber 1999] for this 2-Node chain. The following graph contains the transition probabilities as well as the steady state probabilities together with the measured (or derived) values:



The following estimates are obvious:

$$p_1 = E[Y] = P(Y_i = 1) \approx \tfrac{r}{n} \tag{2}$$

$$p_0 = 1 - E[Y] = P(Y_i = 0) \approx \tfrac{v}{n} = \tfrac{n-r}{n} \tag{3}$$

$$p_{10} = P(Y_i = 0 \mid Y_{i-1} = 1) \approx \tfrac{a}{r} \tag{4}$$

$$p_{11} = P(Y_i = 1 \mid Y_{i-1} = 1) \approx \tfrac{b}{r} \approx \tfrac{r-a}{r} \tag{5}$$

$$p_{01} = P(Y_i = 1 \mid Y_{i-1} = 0) \approx \tfrac{c}{v} \approx \tfrac{a}{n-r} \tag{6}$$

$$p_{00} = P(Y_i = 0 \mid Y_{i-1} = 0) \approx \tfrac{d}{v} \approx \tfrac{n-r-a}{n-r} \tag{7}$$

The analysis of the LRE algorithm from [Schreiber 1999] is based on an a-posteriori argument for the distribution of p_1, given the measured values n, r, a: Starting point are the densities $f_{p_{10}}$ and $f_{p_{01}}$ of possible values p_{10} and p_{01} that 'fit' to n, a, r:

$$f_{p_{10}}(x) = (r+1)\binom{r}{a} x^a (1-x)^b \tag{8}$$

$$f_{p_{01}}(x) = (v+1)\binom{v}{c} x^c (1-x)^d \tag{9}$$

Using these densities, it is shown that the set of values p_1 that 'fit' to n, r, a is approximately normal distributed for sufficiently large values of n, r, v, a, b:

$$p_1 \sim N(\frac{r}{n}, \sigma^2) \tag{10}$$

with

$$\sigma^2 \approx \frac{rv}{n^3} \cdot (\frac{2rv}{an} - 1) \tag{11}$$

Instead of defining confidence intervals, the *limited relative error (LRE)* $\frac{\sigma}{r/n}$ is proposed. Because $p_0 + p_1 = 1$, σ also is an error measure for the approximation v/n of p_0. If r/n is significantly larger than v/n, then the relative error $\frac{\sigma}{v/n}$ is bigger than $\frac{\sigma}{r/n}$. In [Görg and Schreiber 1996, Schreiber 1999], the authors suppose to perform a simulation until both values are below 0.05. For the validity of the approximations (that are derived via an application of the central limit theorem), they give only very simple 'large sample conditions': $n > 1000$; $r, v > 100$; $a, b, c, d > 10$.

Estimation of the sojourn times

Unfortunately, the analysis in [Schreiber 1999] is strictly concentrated on the statistics of 2-Node Markov chain and so it does not give hints how good the approximations are in case that condition (1-LRE) does not hold. In the following we present a new analysis of the LRE algorithm, where we will concentrate on confidence intervals instead of the LRE as error measure. Our analysis is better suited to understand the pros and cons of the LRE approach.
The time series $Y_1, Y_2, Y_3, \ldots$ defines two sequences $(G_i^{(1)})$ and $G_i^{(0)}$ of sojourn times in state 1 resp. state 0 such that

$$Y_1Y_2Y_3\ldots = \underbrace{1\ldots1}_{G_1^{(1)}}\overbrace{0\ldots0}^{G_1^{(0)}}\underbrace{1\ldots1}_{G_2^{(1)}}\overbrace{0\ldots0}^{G_2^{(0)}}\ldots$$

if the initial state happens to be $Y_1 = 1$ or

$$Y_1Y_2Y_3\ldots = \overbrace{0\ldots0}^{G_1^{(0)}}\underbrace{1\ldots1}_{G_1^{(1)}}\overbrace{0\ldots0}^{G_2^{(0)}}\underbrace{1\ldots1}_{G_2^{(1)}}\ldots$$

if the initial state is $Y_1 = 0$.
When considering the LRE-algorithm, we see that using the 2-Node Markov chain to model the behavior of the simulation contains three basic assumptions concerning these sojourn times:

- (2-LRE) The sojourn times $(G_j^{(1)})$ for state 1 are supposed to have identical geometrical distribution $G^{(1)}$ given by probability p_{10}.
- (3-LRE) The sojourn times $(G_j^{(0)})$ for state 0 are supposed to have identical geometrical distribution $G^{(0)}$ given by probability p_{01}.
- (4-LRE) The random sequences $(G_i^{(1)})$ and $(G_i^{(0)})$ are supposed to be independent.

From the properties of the geometric distribution and (2-LRE) we get

$$\mu_1 := E[G_j^{(1)}] = 1/p_{10} \tag{12}$$

and

$$\sigma_1^2 := Var(G_j^{(1)}) = p_{11}/p_{10}^2 \tag{13}$$

The same holds for state 0, where (3-LRE) implies

$$\mu_0 := E[G_j^{(0)}] = 1/p_{01} \tag{14}$$

and

$$\sigma_0^2 := Var(G_j^{(0)}) = p_{00}/p_{01}^2 \tag{15}$$

μ_1 and μ_0 are the mean sojourn times, so we have

$$E[Y] = \frac{\mu_1}{\mu_1 + \mu_0} \tag{16}$$

Now suppose that we perform a simulation exactly until the first a samples of sojourn times $g_j^{(1)}$ in state 1 and the first a sojourn times $g_j^{(0)}$ in state 0 are completed. We still measure the same three values n, a, r as above. So in contrast to the previous chapter, where n was fixed and a and r were depending on n, now a is fixed and $n = n(a)$ and $r = r(a)$ are depending on a! We still use $v = n - r$. In fact we have

$$r = \sum_{j=1}^{a} g_j^{(1)} \tag{17}$$

$$v = \sum_{j=1}^{a} g_j^{(0)} \tag{18}$$

Now consider

$$R_a^{(1)} := 1/a \cdot \sum_{j=1}^{a} G_j^{(1)} \tag{19}$$

and

$$R_a^{(0)} := 1/a \cdot \sum_{j=1}^{a} G_j^{(0)} \tag{20}$$

$R_a^{(1)}$ is the mean value of a geometrically distributed random variables, that are independent because of (4-LRE). If a is large enough, we may apply the central limit theorem stating that $R_a^{(1)}$ has approximately normal distribution with mean μ_1 and variance σ_1^2/a.
Our measured value r/a is indeed just a sample $r/a = 1/a \cdot \sum_{j=1}^{a} g_j^{(1)}$ of $R_a^{(1)}$, so we may deduce confidence intervals for μ_1

$$P(\ |\mu_1 - r/a| < z \cdot \sigma_1/\sqrt{a}\ ) \leq \alpha(z) \tag{21}$$

using the well known function $\alpha(z) = norm(z) - norm(-z)$ for confidence levels.
A similar result holds for $R_a^{(0)}$ and μ_0:

$$P(\ |\mu_0 - v/a| < z \cdot \sigma_0/\sqrt{a}\ ) \leq \alpha(z) \tag{22}$$

Using (4-LRE), we see that $R_a^{(1)}$ and $R_a^{(0)}$ are independent. But the fraction

$$\frac{R_a^{(1)}}{R_a^{(1)} + R_a^{(0)}} \tag{23}$$

of two normal distributed random variables is certainly *not* normal distributed! However, this does not contradict the result of [Schreiber 1999] stating the normal distribution of p_1, as we will show in the following:

$$\begin{aligned}
&\frac{R_a^{(1)}}{R_a^{(1)} + R_a^{(0)}} - \frac{\mu_1}{\mu_1 + \mu_0} \\
&= \frac{R_a^{(1)} \cdot (\mu_1 + \mu_0) - \mu_1 \cdot (R_a^{(1)} + R_a^{(0)})}{(R_a^{(1)} + R_a^{(0)}) \cdot (\mu_1 + \mu_0)} \\
&= \frac{R_a^{(1)} \cdot \mu_0 - \mu_1 \cdot R_a^{(0)}}{(R_a^{(1)} + R_a^{(0)}) \cdot (\mu_1 + \mu_0)} \\
&= \frac{(R_a^{(1)} - \mu_1) \cdot \mu_0 - \mu_1 \cdot (R_a^{(0)} - \mu_0)}{(R_a^{(1)} + R_a^{(0)}) \cdot (\mu_1 + \mu_0)} \\
&\approx \frac{(R_a^{(1)} - \mu_1) \cdot \mu_0 - \mu_1 \cdot (R_a^{(0)} - \mu_0)}{(\mu_1 + \mu_0) \cdot (\mu_1 + \mu_0)} \\
&= R_a^{(1)} \cdot \frac{\mu_0}{(\mu_1 + \mu_0)^2} - R_a^{(0)} \cdot \frac{\mu_1}{(\mu_1 + \mu_0)^2}
\end{aligned} \tag{24}$$

The relative(!) error of this approximation tends to 0, if $R_a^{(1)}$ approaches μ_1 and $R_a^{(0)}$ approaches μ_0. So the precision of the approximation will increase with a, as the variances σ_1^2/a and σ_0^2/a converge to zero.
In consequence, for large a the distributions must be very similar, as soon as their variances are small compared to μ_1 and μ_0. As the class of normal distributed random variables is closed under finite sums and multiplication with a constant value, (24) is obviously normal distributed. Its mean is 0, and its variance turns out to be

$$\sigma_a^2 := \frac{\sigma_1^2 \mu_0^2 + \sigma_0^2 \mu_1^2}{a \cdot (\mu_1 + \mu_0)^4} \tag{25}$$

Substituting the values from (12) to (15), we see that (25) and (11) are identical w.r.t. the approximations (4) to (7):

$$\sigma_a^2 \approx \frac{rv}{n^3} \cdot (\frac{2rv}{an} - 1) \approx \sigma^2$$

Putting all things together, we get the following summary:

- For large a, (23) is approximately(!) normal distributed with mean $E[Y]$ and variance (25), and we have the sample r/n of (23):

$$r/n = \frac{r/a}{n/a} = \frac{r/a}{r/a + v/a}$$

- Our analysis leads to the same results as [Schreiber 1999], but now r/n turns out to be a sample from a normal distribution, while in the original paper the variance of p_1 was constructed in a fairly complicated way with an a-posteriori argument from n, a, r.

- It is now clear to see where large sample conditions are necessary: a should allow application of the the central limit theorem:

$$a > 10 \tag{26}$$

(which should be sufficient, as here we have geometric distributions), and the approximation (24) should be sufficiently precise:

$$\sigma_1/\sqrt{a} << \mu_1 \quad , \quad \sigma_0/\sqrt{a} << \mu_0 \tag{27}$$

which is equivalent to

$$1 >> \frac{(r-a)n}{a^3 r} \quad , \quad 1 >> \frac{(v-a)n}{a^3 v} \tag{28}$$

Confidence intervals for LRE

Although the a-posteriori analysis from [Schreiber 1999] results in a normal distribution, the author of [Schreiber 1999] did not introduce confidence intervals, perhaps because the proof did not prove r/n to be a sample!
But with our approach, it is very natural to construct confidence intervals. In addition, we are now able to compare the results of the LRE with the batch means method.
Using the well known relation $\alpha(z) = norm(z) - norm(-z)$ for confidence levels we get

$$\begin{aligned} \alpha(z) &= P(-z \le (E[Y] - r/n)/\sigma \le z) \\ &= P(\ |E[Y] - r/n| \le z \cdot \sigma\ ) \end{aligned}$$

The values $\alpha(1.64) \approx 0.9$, $\alpha(1.96) \approx 0.95$ and $\alpha(2.58) \approx 0.99$ are frequently used to get 90%, 95% or 99% of confidence. As $\alpha(1) \approx 0.68$, using σ as error measure corresponds to a confidence level of only 68%!
So according to the suggestions in [Schreiber 1999] cited above, a simulation should already be stopped (as being precise enough), as soon as the 68% confidence interval has a radius of 5% of the mean value. At that time, the more usual 95% confidence interval has a radius of about $1.96 \cdot 5\% = 9.75\%$ of the mean value! We do not think that this precision is high enough for applications and suggest that simulations should be performed without this stopping criterium.

Discussion and experimental results

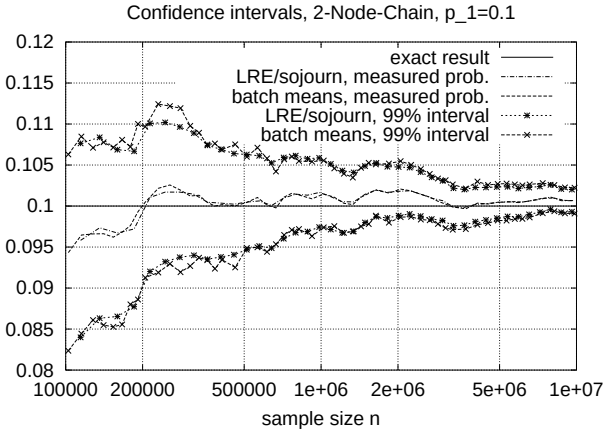
Obviously, condition (1-LRE) will not be true in most real simulation scenarios. On the other hand, the one step transition probabilities lead to correct values for the LAG-1-autocorrelation of the series (Y_i). Unless additional data is measured from the simulation, the assumption of Markovian behaviour implies estimates on the higher type autocorrelations that can hardly be improved.

Our sojourn time analysis (25) shows that there is a big influcence of the sojourn time variance on the resulting confidence intervals. If the real variances are smaller than σ_1 resp. σ_0, the LRE-algorithm will result in unneccessarily large intervals. On the other hand, the intervals are too small if the real sojourn time variances are larger than σ_1 resp. σ_0.

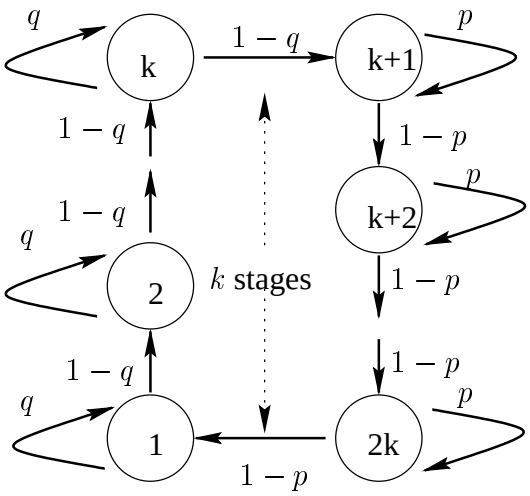
In the following we present the results of a few tests to illustrate this behaviour. Using three different types of time series (Y_i) and for a range of samples from 10^5 to 10^7 we measured 99% confidence intervals for the LRE algorithm and for an implementation of the Law-Carson algorithm for the batch means method (see e.g. [Bratley et al. 1987]). The examples were constructed from several discrete Markov chains (X_i) with $Y_i = 1 \Leftrightarrow X_i \in G$, where the sets G were chosen subsets of the state spaces.

The intervals due to the batch mean method do not rely on the LRE hypotheses, so they vary correctly with the different variances of the sojourn times.

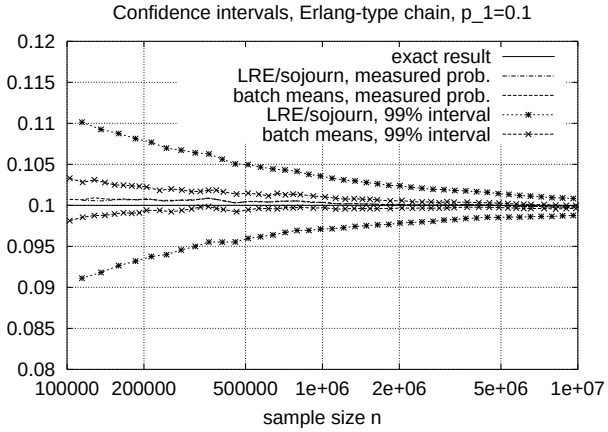
We first start with a setting that fits to (1-LRE), i.e. the time series (Y_i) indeed came from a 2-Node Markov chain with $Y_i = X_i$. We chose $p_1 = 0.1$ and $p_{10} = 0.06$. Please note that the other transition probabilities are uniquely defined by p_1 and p_{10}. As expected, the intervals are almost identical:



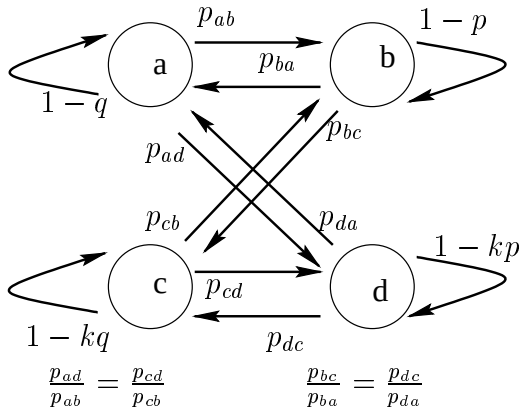
The second test used a series with sojourn times of significant smaller variance. It is based on the following 2k-Node Markov chain with $G = \{1, ..., k\}$, i.e. $Y_i = 1 \Leftrightarrow X_i \le k$.



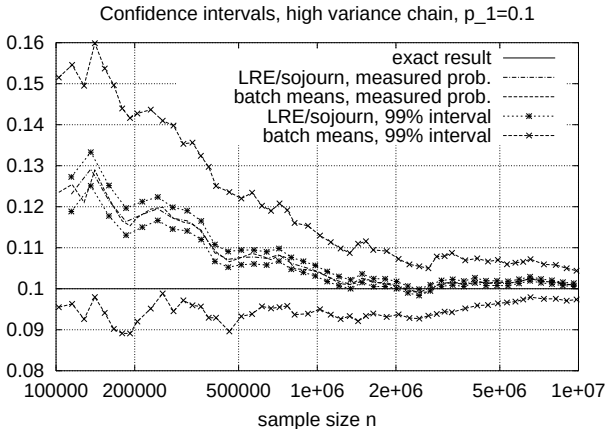
We tried this example with $k = 10$ and $p = 0.99$. As expected, the confidence intervals are much larger than for the batch means method:



The third test used a 4-Node chain with highly variant sojourn times, where $G = \{a, c\}$:



We tried an example with $k = 100$, so the sojourn times are build by mixtures of two geometric distributions differing by a factor of 100. As expected, the confidence intervals are way too small in this case:



Please note that in all examples, the assumption (4-LRE) about the independency of the sojourn times is valid! This is true also for any Markovian birth-death process, so (2-LRE) and (3-LRE) seem to be more critical than (4-LRE).

Conclusion

From the examples, the disadvantages of the 2-Node approach can be clearly seen. On the other hand, this approach can be easily incorporated into rare event simulations, see e.g. [Görg and Schreiber 1996]. Here the batch means method is not applicable at all. So regardless of these above problems, the LRE algorithm is very useful in this field.

Our new analysis shows where improvements are possible: At the time, we are working on an algorithm measuring σ_1 and σ_0 directly from the times series instead of using the conditions (2-LRE) and (3-LRE). Initial results on the gained confidence intervals for non-geometric sojourn times are very promising.

REFERENCES

[Bratley et al. 1987] Bratley, Paul; Fox, Bennett L.; and Schrage, Linus E. 1987. *A Guide to Simulation.* Springer, New York

[Görg and Schreiber 1996] Görg, C.; Schreiber, F. "The RESTART/LRE Method for Rare Event Simulation." In *Proceedings of the 1996 Winter Simulation Conference.* Coronado, California USA, pp390-397.

[Schreiber 1984] Schreiber, F. "Time Efficient Simulation: The LRE-algorithm for producing empirical distribution functions with limited relative error." In *AEÜ, 38, 1994*, pp93-98.

[Schreiber 1999] Schreiber, F. "Reliable Evaluation of Simulation Output Data: A simplified Formula Basis for the LRE-Algorithm" In Proceedings of the *MMB '99, Trier.* VDE Verlag Berlin, ISBN 3-8007-2472-3, pp 137-152.

Many additional references can be found at the following web page:

`www.comnets.rwth-aachen.de/rare-event/`