Some Observations on Importance Sampling and RESTART

— preliminary paper for RESIM2006 —

Pieter-Tjerk de Boer*

Abstract

Two well-known techniques for rare-event simulation of queueing models are compared with respect to their achievable performance, and difference in robustness; furthermore, a theoretical mapping between both techniques is devised and explored with the aim of obtaining a better understanding of both methods' operation and performance.

1 Introduction

Over the last two decades, there has been a considerable interest in techniques for efficient estimation of rare-event probabilities in queueing models; to a large extent, this interest derives from applications in modelling of telecommunication networks. Two methods have received much attention: *Importance Sampling*, and *Splitting* or *RESTART* (the latter two being minor variations of the same principle, see [VAVA06]; henceforth, RESTART will be considered).

On each of the two methods, many research papers have been written. Each of the two methods has been shown to perform well in some situations; also, for each of them more complicated problems are known where they do not work well, or require more effort and insight on the part of the simulation user. So far, not much has been written about relationships and comparisons between the two methods, apart from a few remarks such as in Section 2.5 of [Gar00], in [Add02], and Section 3.5.3 in [dBN02]. The present paper makes an attempt to explore such relationships in more detail.

In Section 2, both methods will be briefly introduced, and some observations (mostly from literature) about their efficiencies are listed. Section 3 builds on this, to give an intuitive motivation for the observed better robustness of RESTART. In Section 4, a suggestion for a way of mapping IS onto RESTART and vice versa is introduced, and its consequences explored. Finally, Section 5 summarizes the findings.

Purposely, this paper includes almost no mathematics, and as a consequence is not very precise in many aspects. For example, details such as minor inefficiencies due to discreteness of the state space are ignored. This is done for simplicity, to emphasize the main ideas; in fact, many of the ideas are rather preliminary (some are less than a week old at the time of writing), and there are no solid conclusions yet. As such, the paper should mostly serve as a starting point for further thoughts and discussions at the RESIM2006 workshop.

2 The methods and their potential efficiencies

In this section, both methods are briefly introduced and their potential efficiencies are discussed, drawing heavily from literature. But first some terminology needs to be introduced.

Throughout, we will consider the estimation of transient probabilities of the following type: the probability of reaching a (rare) set (called the "target"), starting from some given state and before returning to that state. In a queueing context, the starting state could be the empty system, and the target state(s) could be those in which a queue is full.

The "relative error" (RE) is defined as the estimator's standard deviation divided by its mean. It is a useful measure for judging a simulation scheme's efficiency in the limit of the target probability going to zero, e.g. as the buffer size of a queueing model goes to infinity.

If the target probability goes to zero exponentially fast in a model parameter like the buffer size, and the simulation method is such that the effort required for estimating it with a fixed RE increases less than exponentially fast, the method is said to be asymptotically efficient; equivalently, for a given cost the RE increases less than exponentially fast. For the "cost", several measures may be used, such as the number of sample paths needed, or (an estimate of) the computational effort.

* Dept. of Electrical Eng., Mathematics and Computer Science, University of Twente, Enschede, The Netherlands; ptdeboer@cs.utwente.nl

2.1 Importance Sampling

The basic principle of Importance Sampling (henceforth abbreviated to IS) is that the underlying probabilities of the model are changed in such a way as to make the (rare) event of interest occur more frequently during the simulation; this change is called a “change of measure” or “tilting”. In order to still obtain an unbiased estimate of the target event’s probability, a *Likelihood Ratio* (henceforth called LR) is calculated, representing the ratio between the probability of the sample path under the original probability distributions, and the modified (“tilted”) probability distributions. Thus, each observation of the target event can be weighted appropriately. An overview of IS can be found in [Hei95].

The main difficulty in applying IS is deciding how to modify the model’s underlying probability distributions; this greatly affects the resulting variance. In practice, an important distinction has been found to be between *state-independent* and *state-dependent* changes of measure. The former just replace some parameters (arrival rates etc. in case of a queueing model) of the model without regard to the system state. The latter do allow such dependence, which may be expected to improve the simulation performance if done properly; however, because of its many degrees of freedom a state-dependent tilting may also be quite difficult to choose properly [KN02, dB00, dBN02, dB06].

Efficiency

With a proper (state-dependent) change of measure, IS can give *zero variance*. This happens when the change of measure is such that all sample paths reach the target, and they do so with precisely the same likelihood ratio, which then is equal to the target probability. However, finding this change of measure is at least as hard as calculating the probability of interest itself.

On the other hand, with a very improper change of measure, the variance can be *infinite*. This happens when there is no upper bound on the likelihood ratio on possible sample paths, and sample paths with large LRs are (in a precise sense) likely enough in the tilted system (cf. Lemma 3.2 in [RJ04]). If the LR is upper-bounded, infinite variance obviously cannot occur.

Even when the variance is not infinite, a problem known as *underestimation* may occur. Roughly speaking, this may happen when some sample paths to the target event are “too” unlikely under the tilting; consequently, they have a high LR. On the one hand this causes a large variance estimate if they do occur during the simulation run, but on the other hand it causes underestimation if they do not occur (which is quite likely in a finite simulation run). See the “underestimation theorem” in [DT93].

For many problems, an *asymptotically efficient* change of measure can be found. For simple problems (such as a single queue, see [Sad91]) even a state-independent change of measure may be asymptotically efficient, but for more complex problems this typically does not suffice [dB06]. In queueing models for which an asymptotically efficient IS scheme is known, typically the relative error for a fixed number of sample paths becomes constant (“bounded relative error”), although in some cases a linear growth with the overflow level has been found; see e.g. [Sad91, GK95]. The computational *effort* per sample path is not often discussed in the literature. For a first estimate, we may assume that this effort is proportional to the number of steps on that sample path, which in turn may be assumed to be proportional to the overflow level. Thus, the effort per sample path tends to be proportional to the overflow level; in case of bounded relative error, we thus find that the effort required for a constant relative error with IS simulation increases *linearly* with the overflow level.

2.2 RESTART

With RESTART simulation, the number of observations of the target event is artificially increased by “splitting” any “promising” sample path: if a sample path is “near” the target state by some measure, its state is stored and multiple copies of it are simulated henceforth. This splitting can actually be done several times on the way from start to target state [VA$^+$94].

For a given (queueing) problem, a RESTART simulation scheme is defined by the number of offspring (denoted R) at each splitting, and the sets of states in which the splitting occurs. The latter are typically determined by a so-called “importance function” (henceforth IF): a function which assigns a scalar to every state, and whose levelsets form the sets of states at which splitting occurs. The final estimator is then simply the number of times the target was observed, divided by the number of sample paths started, and divided by the total splitting factor r, which is the product of the splitting factors R at each level. The total splitting factor thus acts as a weighting factor $1/r$; in contrast to the weighting factor LR in IS, this weighting factor in RESTART is the same for all trajectories. The right-hand side of Figure 1 gives an example with $R = 3$.

Given an importance function, choosing the right levels at which to split and the right splitting factor R is not too hard, and there is even some margin without losing asymptotic efficiency [VAVA06]. The hard part is choosing the importance function. In queueing models, the definition of the target probability can often inspire a simple choice, e.g., the level of the queue whose overflow probability one tries to estimate. However, this simple choice does not always work well [GHSZ98]; then the IF needs to depend also on other state variables, making its choice less trivial. (This need seems to be the RESTART counterpart of needing state-dependence in IS.)

Efficiency

Fundamentally, RESTART cannot have zero variance. This is because there are always both sample paths that do reach the target event, and ones that don't, so the final estimator is an average of zero and non-zero numbers.

Furthermore, RESTART cannot have infinite variance. This is because every sample path contributes a bounded amount (namely either 0 or $1/r$) to the final estimator.

For many queueing problems, *asymptotically efficient* RESTART schemes exist. In the context of RESTART, asymptotical efficiency is typically defined in terms of computational effort, and not number of sample paths, since the latter is not clearly defined due to the splitting. Asymptotically efficient RESTART schemes typically require an effort that is proportional to the square of the log of the target probability [VA+94], [Gar00, p. 20], [VAVA06, eq. (13)]; i.e., in queueing models with exponentially decaying overflow probability, such an effort is proportional to the square of the overflow level.

The latter can be understood intuitively. One can consider a RESTART simulation as a set of k independent estimations of the probability of reaching the $(i+1)$th level from the ith level. Optimally, these level-to-level probabilities are the same at every level and equal to the inverse of the number of offspring [VA+94, Gar00]. Thus, as the overflow level is doubled, typically the number of splitting levels k is also doubled. Since the final estimate is the product of the k level-to-level estimates, and assuming independence, its RE equals the RE of the individual estimates times $\sqrt{k}$. To keep this constant as k doubles, each of the estimates' REs must be improved by a factor of $\sqrt{2}$, which requires a factor 2 more effort for each estimation. Furthermore, recall that the total number k of such estimations has also doubled, so the total effort quadruples as the overflow level doubles.

RESTART can also suffer from *underestimation* when the importance function is chosen improperly; see a numerical example in [GHSZ98].

2.3 Comparison

The following table summarizes the findings regarding possible efficiencies so far:

	Importance Sampling	RESTART
Zero variance	✓	
Asymptotic efficiency (effort growth rate, at best)	✓ (linear)	✓ (quadratic)
Underestimation	✓	✓
Infinite variance	✓	

Note that IS has the potential to be much *better* than RESTART. It can achieve zero variance; admittedly, this is unachievable in practice, but the zero variance possibility does play a role in some schemes that try to adaptively converge to (near) it [dBN02, ABJ06]. More practically relevant is the fact that although both methods can achieve asymptotic efficiency, IS' asymptotic efficiency may be better than RESTART's.

Note also that IS has the potential to do much *worse* than RESTART: IS can have infinite variance, RESTART cannot. On the one hand, this is a strong point in favour of RESTART. On the other hand, both methods can suffer from underestimation, which actually is caused by a very large (though possibly finite) variance. The difference between infinite and finite-but-large variance may be of theoretical interest, but not so much of practical relevance.

In conclusion: in terms of range of achievable performance, IS seems to be more general than RESTART (but this should *not* be taken as implying that RESTART is a special case of IS).

3 Robustness – an observation

RESTART cannot have infinite variance, while IS can, which makes RESTART somewhat more robust. Let's take a closer look at this.

As stated before, infinite variance in IS is caused by sample paths whose weight (i.e., LR) can be unboundedly large. If any single step in an IS simulation only makes a finite contribution to the LR, and the state space is finite, such an unbounded growth must be associated with cycles in the sample paths: for every time the cycle is passed through, the LR may increase by some factor, so it can grow unboundedly. This is corroborated by Lemma 3.2 in [RJ04], which gives a sufficient condition for infinite variance relating to cycles in sample paths.

Now, let us consider cycles in sample paths in RESTART. If such a cycle does not cross one of the splitting levelsets (i.e., is contained entirely between two such sets), there is obviously no difference between the first time a state on the cycle was reached and any later time. If the cycle does cross a splitting levelset, it must cross it equally often in the downward and in the upward direction. For every up-crossing, the path is split into multiple “offspring”. However, for every down-crossing, only one of those offspring continues, the rest are stopped. As a consequence, the number of offspring on the cycle does not change, and thus the path's weight stays constant, no matter how often the cycle is gone through.

Thus, we see a fundamental difference. RESTART by its nature binds the weight of paths, and whatever happens to paths, to the states they are in. The RESTART simulator does not add any extra state variables beyond those in the model. On the other hand, in IS the history of the path may influence the state of the simulator: even if the modelled system is in the same state, one of the simulator's state variables, namely the LR, may differ. This intuitively indeed “looks” like it may introduce extra variance.

Removing this extra, “hidden” state variable would mean that on two visits to the same system state, the LR should be the same. In other words, the LR must be a function of the state only. Indeed, the zero-variance change of measure has this property. (The importance of the LR depending only on the state is further explored in [Jun01].) Unfortunately however, it is not easy to construct a change of measure with this property, see Section 4.4. Note further that dependence of the LR on only the state by itself is not a guarantee for efficient simulation: the standard (untilted) simulation has an LR which is constant (namely, equal to 1) in any state, but does not give an efficient simulation.

This fundamental difference in robustness seems to come from the way in which the simulation is modified. In RESTART, this modification (i.e., splitting and dropping of sample paths) is strongly tied to the system *state*, causing the weight of any sample path to relate directly to its state. In IS however, the modification is tied to state *transitions*; of course, the current model state also depends on the transitions that have been made, but there is no natural guarantee that the LR and the model state remain tied together. One could also say that in IS the modification is specified more *indirectly* than in RESTART.

4 Mapping

In this section, we explore a possible connection between the two methods. We start by mentioning some properties of zero-variance IS, and its relation to the optimal importance function for RESTART. Next, we try to use this relationship to map any RESTART simulation to a corresponding IS simulation, and vice versa.

4.1 IS with zero variance

It is well-known (see, e.g., [Jun01, dBN02]) that the zero-variance change-of-measure has the following properties:

- The likelihood-ratio at any point on a sample path is a function of only the state.
- As a consequence: any cycle in a sample path has a likelihood ratio contribution[1] of 1.
- The likelihood-ratio at any state S is equal to the probability of reaching the target from the starting state, divided by the probability of doing so from the present state S.
  Thus, LR starts at 1 in the starting state, and along typical sample paths decreases, until it equals the probability of interest in the target state.

4.2 Optimal RESTART

According to Theorem 2 in Section 5.2 of [Gar00], and others, the optimal importance function is such that its levelsets are the sets of states from which the probability of reaching the target is equal. Clearly, the likelihood ratio from zero-variance IS satisfies this condition, and thus could be used as an optimal IF for RESTART (as already noted in Section 3.5.3 of [dBN02]).

Furthermore, for optimal RESTART simulation the splitting factor R at each level should be the inverse of the probability of reaching the next level (this makes sense intuitively; otherwise, the number of branches may either explode or go to zero as the target level is approached; [VA$^+$94, Gar00]). As a consequence, given two consecutive levels chosen as levelsets of the IS likelihood ratio, the splitting factor should be equal to the likelihood ratio of the lower divided by the likelihood ratio of the higher level. This is illustrated in Figure 1. (In this figure, and in the sequel, a splitting factor $R = 3$ is assumed for simplicity.)

Note that from a practical point of view, it doesn't make sense to first painstakingly derive the zero-variance change of measure, and then use that to provide an importance function for RESTART simulation, since the latter would surely have a non-zero variance. However, we can try to use this same principle to create a mapping between IS changes of measure and RESTART importance functions.

[1]Note that the total LR is the product of the LR of individual transitions, so a contribution of 1 means the LR stays the same.

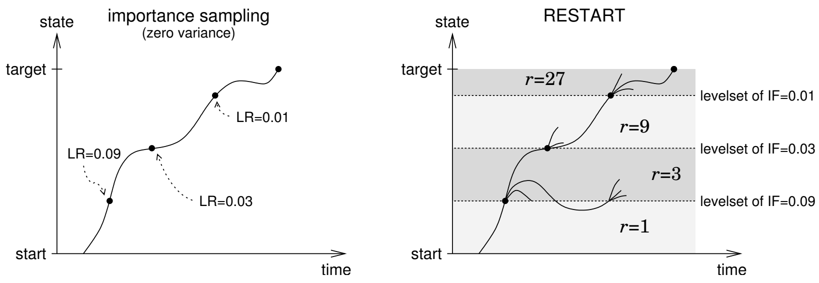

Figure 1: Mapping of zero-variance IS onto RESTART

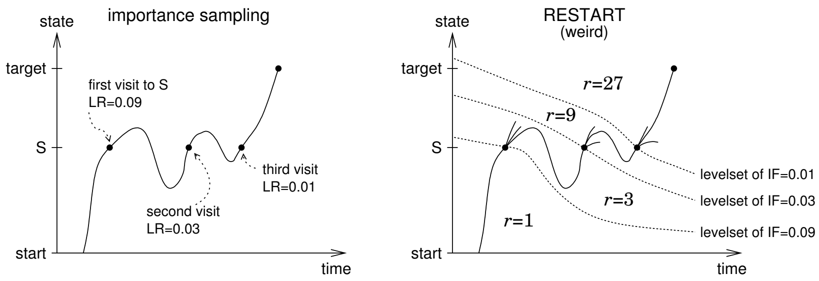

Figure 2: Mapping of IS with $\mathrm{LR}_{\mathrm{cycle}} = \frac{1}{3}$ onto RESTART

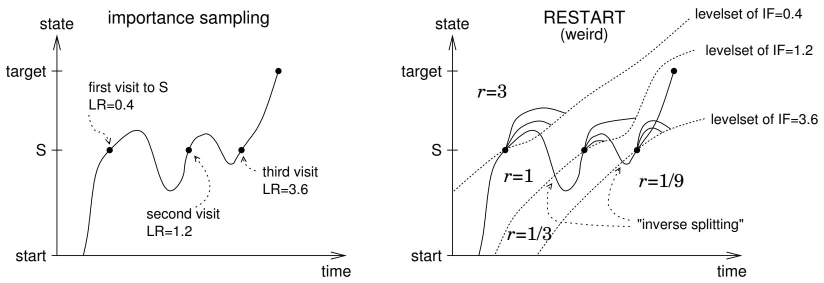

Figure 3: Mapping of IS with $\mathrm{LR}_{\mathrm{cycle}} = 3$ onto RESTART

4.3 Mapping IS onto RESTART

Let us consider the situation where we already have an importance sampling estimator, and thus know (can calculate) its likelihood ratio at any point on any given sample path. Can we construct a corresponding importance function for RESTART?

If the change of measure is such that the LR is a function of only the state, the mapping is simple: just use the LR as the importance function. However, only very few changes of measure have this property.

Otherwise, the mapping is not so straightforward. It is still feasible though, if we allow the importance function to not just depend on the state, but also on the history of the sample path, just like the LR does in this case. This means that a state which at some time is at a specific level of the importance function, may be at a different level at a later time on the same sample path. In order to illustrate this, we look at sample paths that contain cycles.

Let us first consider the case where in a cycle, the likelihood ratio *decreases*; so the second time a state is reached, the LR is *lower*, which implies that it is at level of the IF that originally corresponded to a *rarer* set of states. See Figure 2 for an example where the LR contribution of the cycle under consideration is 1/3.

Note that in these figures, the vertical axis represents some state variable (e.g., queue length). In Figure 1, the vertical axis could *also* be interpreted as the IF axis, but in Figure 2 this is no longer possible since the IF now depends on the path history; therefore, IF levelsets are indicated by dotted lines. These IF levelsets should not be taken too literally, though: the IF does not *explicitly* depend on time; the figure is only meant to illustrate that upon returning to a state at a later time, that state may be at a different IF level, which may cause splitting to occur again.

Cycles with LR<1 in IS are rather benign, they don't cause the LR to explode and thus don't cause infinite variance. The RESTART counterpart does look a bit “risky”: with the levelsets decreasing in time, more and more levelsets appear between state S and the target state, so the expected number of splittings (and thus the effort) might go to infinity. This would happen if $p \cdot R > 1$, where p is the probability of the cycle from S to S; however, since $R = 1/LR_{\text{cycle}}$, we would need $p/LR_{\text{cycle}} > 1$; with $LR_{\text{cycle}} = p/p'$ where p' denotes the tilted probability of the cycle, we find $p' > 1$, so this would correspond to an *impossible* change of measure. Thus, any feasible IS change of measure will never be mapped onto such an infinite-effort RESTART.

Secondly, let's consider the case where in a cycle the LR *increases*. This is potentially risky, as noted before, possibly leading to infinite variance in IS. This case is illustrated in Figure 3. Our choice $R = 3$ implies that the levelsets of the IF (= LR, in this mapping) at which splitting happens are at levels that also differ by a factor of 3; in the example, they are at 0.4, 1.2 and 3.6. Note that now the latter two levelsets are *down*crossed before they are *up*crossed.

Normally in RESTART, at an upcrossing a sample path is split in (e.g.) 3 branches, and at the next down-crossing only one of those 3 branches survives. However, if there is a downcrossing without a previous upcrossing, there are no 3 branches from which one can be allowed to survive; the closest alternative then is to let one out of every 3 sample paths survive, at random or deterministically. (This happens at the points marked “inverse splitting” in the figure.) Recall that in the end, in RESTART any trajectory that hits the target state is weighted by the inverse of the total splitting factor up to that point; this is $1/r$, where r multiplied by 3 at every upcrossing, and divided by 3 at every downcrossing. To be consistent, this must also be done at each downcrossing below the initial level; as illustrated in the figure, this leads to $r < 1$, and thus to weights $1/r > 1$.

Now, if the splitting factor is large, and the probability of this cycle is large too, then the weight of typical trajectories to the target will be very large; at the same time, only few trajectories will survive the “inverse splitting”, so the expectation can still be correct; the variance however will be large.

Just like in its IS counterpart, this can lead to infinite variance. To see this, note that the contribution of these paths to the second moment of the estimator is $\sum_{i=1}^{\infty}(p/R)^i R^{-2i}$, where p is the probability of the cycle and thus p/R is the probability of the cycle in this RESTART scheme (the $1/R$ accounts for the inverse splitting); and R^{-2i} is the square (because we calculate the second moment) of the weight of a sample path having made i cycles. Substituting $R = 1/LR_{\text{cycle}}$, one sees that this second moment goes to infinity under the same condition as the condition for infinite variance in IS from Lemma 3.2 in [RJ04].

Note that this way of mapping IS onto RESTART typically leads to a “weird” RESTART, in the sense that its IF depends on the path history. It does not make any sense at all to actually perform a simulation in this way; the only purpose of this theoretical mapping may be to help understand the two methods.

4.4 Mapping RESTART onto IS

Now, let's try the reverse. We somehow have a RESTART setting, that is, we have levelsets, labeled $1 \ldots k$, of an importance function, and we want to construct the corresponding IS simulation. Given levelset i of the importance function, in principle the probability of reaching the target from that levelset can be computed; let's denote it by $P_{A/i}$ (borrowing notation from [VAVA06]). Denoting he probability of reaching the target from the starting state by γ, we can assign to each levelset i the likelihood ratio $LR_i = \gamma/P_{A/i}$. Then our task is to construct an IS change of measure such that in any state, the likelihood ratio is LR_i.

Note that if this would be possible, we would have arrive at a change of measure with the property that the LR is a function of only the state, which is desirable as noted in Section 3.

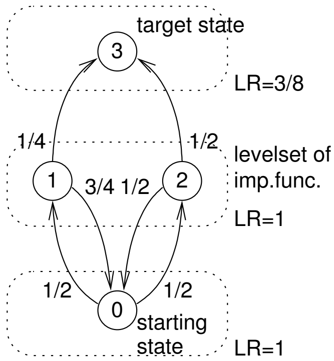

Figure 4: Counterexample for mapping RESTART onto IS

Unfortunately, this is typically not possible. One can see this by considering the trivial example shown in Figure 4. The figure shows a Markov chain with four states, that have been assigned to three levelsets (see dotted lines) by some (non-optimal) importance function. One easily calculates that the probability of reaching the target state 3 from state 0 before returning to state 0 is γ=3/8; furthermore, the probability doing this starting from state 1 (appropriately averaged over states 1 and 2) is also 3/8. Hence, when mapping this to IS, the LR on each of the levelsets should be as indicated in the figure. This implies however that the tilted transition probability from state 2 to state 3 should be (1/2)/(3/8) = 4/3, which is impossible. (This example can easily be generalized.)

Just like we needed to generalize RESTART for the mapping as discussed in Section 4.3 by allowing the importance function to depend on the path history, perhaps we can also generalize IS to make the mapping from RESTART possible. However, at the time of writing the author has not yet found a way to do this.

4.5 Interpretation

The proposed mapping between IS and RESTART is only partially succesful. In the cases where it applies, it seems to give sensible results: good IS is mapped to good RESTART, bad IS is mapped to bad RESTART, and a hypothetical RESTART with infinite effort is mapped to an impossible IS. Unfortunately, the mapping is far from complete.

Going from IS to RESTART, the resulting RESTART is a rather generalized one, allowing the IF levelsets to depend on the path history. The need for this generalization is to be expected, since we noted before that IS allows a larger range of performance than (non-generalized) RESTART.

Going from RESTART to IS, the mapping is usually impossible; this is both somewhat surprising and disappointing. Surprising because IS seemed more general when judged by its performance range. And disappointing, because a mapping in this direction might give insight on how to combine RESTART's robustness with IS.

5 Conclusions

- In terms of performance, IS has the wider range of possibilities: it can give zero variance, infinite variance, and anything in between; RESTART does not "offer" either of the extremes, which makes it both less promising and less risky.
  The effort growth rate in case of asymptotical efficiency is often lower for IS than for RESTART.
- RESTART's extra robustness (i.e., no infinite variance) seems to come from the fact that its modification to the system is tied more directly to the system state; in contrast, in IS the modification is tied to the state *changes*, so the modification is only indirectly tied to the system state.
- A theoretical mapping has been proposed between IS and RESTART, but its success is rather limited.
  Further research into such mappings may be of interest, for better understanding of the methods, and to perhaps combine both methods' advantages.

References

[ABJ06] T. P. I. Ahamed, V. S. Borkar, and S. Juneja. Adaptive importance sampling technique for Markov chains using stochastic approximation. To appear in *Operations Research*, 2006.

[Add02] R.G. Addie. Quantum simulation – rare event simulation by means of cloning, thinning and distorsion. *European Transactions on Telecommunications*, 13(4):387–397, 2002.

[dB00] Pieter-Tjerk de Boer. *Analysis and efficient simulation of queueing models of telecommunication systems*. PhD thesis, University of Twente, 2000.

[dB06] Pieter-Tjerk de Boer. Analysis of state-independent IS measures for the two-node tandem queue. *ACM Transactions on Modeling and Computer Simulation*, 16:225–250, 2006.

[dBN02] Pieter-Tjerk de Boer and Victor F. Nicola. Adaptive state-dependent importance sampling simulation of markovian queueing networks. *European Transactions on Telecommunications*, 13(4):303–315, 2002.

[DT93] Michael Devetsikiotis and J. Keith Townsend. An algorithmic approach to the optimization of importance sampling parameters in digital communication system simulation. *IEEE Transactions on Communications*, 41:1464–1473, 1993.

[Gar00] Marnix J. J. Garvels. *The splitting method in rare event simulation*. PhD thesis, University of Twente, 2000.

[GHSZ98] Paul Glasserman, Philip Heidelberger, Perwez Shahabuddin, and Tim Zajic. A large deviations perspective on the efficiency of multilevel splitting. *IEEE Transactions on Automatic Control*, 43:1666–1679, 1998.

[GK95] Paul Glasserman and Shing-Gang Kou. Analysis of an importance sampling estimator for tandem queues. *ACM Transactions on Modeling and Computer Simulation*, 5(1):22–42, January 1995.

[Hei95] P. Heidelberger. Fast simulation of rare events in queueing and reliability models. *ACM Transactions on Modeling and Computer Simulation*, 5:43–85, 1995.

[Jun01] Sandeep Juneja. Importance sampling and the cyclic approach. *Operations Research*, 49(6):900–912, 2001.

[KN02] D. P. Kroese and V. F. Nicola. Efficient simulation of a tandem jackson network. *ACM Transactions on Modeling and Computer Simulation*, 12(2):119–141, 2002.

[RJ04] R. S. Randhawa and S. Juneja. Combining importance sampling and temporal difference control variates to simulate markov chains. *ACM Transactions on Modeling and Computer Simulation*, 14(1):1–30, 2004.

[Sad91] J. S. Sadowsky. Large deviations theory and efficient simulation of excessive backlogs in a $GI/GI/m$ queue. *IEEE Transactions on Automatic Control*, 36:1383–1394, 1991.

[VA+94] M. Villén-Altamirano, A. Martínez-Marrón, J. Gamo, and F. Fernández-Cuesta. Enhancement of the accelerated simulation method RESTART by considering multiple thresholds. In J. Labetoulle and J. W. Roberts, editors, *Proceedings of the 14th International Teletraffic Congress*, pages 797–810. Elsevier Science, 1994.

[VAVA06] Manuel Villén-Altamirano and José Villén-Altamirano. On the efficiency of restart for multidimensional state systems. *ACM Transactions on Modeling and Computer Simulation*, 16(3):251–279, 2006.