Accident Risk Assessment and Monte Carlo Simulation Methods

Pascal Lezaud, Jaroslav Krystul, Henk Blom

HYBRIDGE

Distributed Control and Stochastic Analysis of Hybrid Systems
Supporting Safety Critical Real-Time Systems Design

WP8: Accident risk decomposition

Accident Risk Assessment and Monte Carlo Simulation Methods

Pascal Lezaud[1], Jaroslav Krystul[2], Henk Blom[3]

9th September 2003

Version: *0.4*
Task number: *8.2*
Deliverable number: *D8.2*
Contract: *IST-2001-32460 of European Commission*

[1] Centre d'Etudes de la Navigation Aérienne, France

[2] Twente University, NL

[3] NLR, NL

DOCUMENT CONTROL SHEET

Title of document: *Accident Risk Assessment Methods*

Authors of document: *P.Lezaud, J.Krystul, H.Blom*

Deliverable number: *D8.2*

Contract: *IST-2001-32460 of European Commission*

Project: *Distributed Control and Stochastic Analysis of Hybrid Systems Supporting Safety Critical Real-Time Systems Design (HYBRIDGE)*

DOCUMENT CHANGE LOG

Version #	Issue Date	Sections affected	Relevant information
0.1	29 Jun. 2003	All	1st draft
0.2	15 Aug. 2003	1, 5, A.	2nd draft
0.3	26 Aug. 2003	1, 2, 5, 6.	3rd draft
0.4	9 Sept. 2003	3, 4.	4th draft

Version 0.4		Organisation	Signature/Date
Authors	Pascal Lezaud	CENA	
	Jaroslav Krystul	TWEN	
	Henk Blom	NLR	
Internal reviewers			

Contents

1 Introduction

Within the HYBRIDGE project, the objective of WP8 is to develop novel methods for the decomposition of risk such that extreme low risk values can be assessed through a hierarchy of conditional Monte Carlo simulations which are scalable relative to the increasing complexity of the application considered. The research within WP8 is organized in four tasks:

8.1 - Review existing risk decomposition and assessment methods, both analytical ones, Monte Carlo simulation approaches and combinations of these two. This review should distinguish between theory based methods and heuristic methods.

8.2 - Develop new risk decomposition and assessment methods. One of the key directions to be explored is the development of risk decomposition methods that make use of the fact that for strong Markov processes the Markov property holds true for stopping times.

8.3 - Development of Conditional Monte Carlo simulation techniques for accident risk assessment that make use of the risk decomposition developed in Task 8.2, and comparison of the new approach with the existing ones identified in Task 8.1.

8.4. - Extend the risk decomposition approach with a recursive Bayesian estimation approach which enables the updating of the accident risk assessment while more and new information is becoming available.

The results of task 8.1 have been reported in [Krys03a]. The current report addresses the results obtained during the performance of Task 8.2.

Accident risk assessment has been widely studied for various safety-critical operations, such as the nuclear and chemical industries, advanced air traffic management (ATM) and many other. All accident risk assessment approaches can be subdivided into two groups: approaches based on statistical analysis of collected data and those that are based on the modelling of the processes leading to the accident. The statistical analysis of extreme values needs a long observation time since the very low probability of the events considered. Statistical approaches are based on the standard extreme value distributions as the Gumbel, Fréchet and Weibull laws [Reiss97]. The modelling approaches firstly consist in formalizing the system considered and secondly by using mathematical or simulation tools obtaining some estimate. The aim of work package WP8 is to study the second group of accident risk assessment methods and to develop novel methods for the obtaining accurate estimates of rare event probabilities.

We distinguish three different types of modelling approaches: analytical, numerical and simulation techniques. Analytical, numerical and standard Monte Carlo approaches have been studied in Task 8.1 of the work package WP8. Analytical and numerical approaches appeared to be useful, but can require many simplifying assumptions (see [Krys03a]). This finding agrees with conclusions obtained from studies of realistic ATM example in D2.2 (see [Blom et al 03a]). This causes Monte Carlo simulation to be a practical alternative when the analysis calls for fewer simplifying assumptions. However, obtaining accurate estimates of rare event probabilities, say about 10^{-9} to 10^{-12}, is not realistic

using straightforward Monte Carlo simulation. Thus, the objective of Task 8.2 is to study effective Monte Carlo simulation speed-up techniques which are complementary to speed-up by parallel or distributed Monte Carlo simulation architectures.

Many techniques for reducing the number of trials in Monte Carlo simulation have been proposed, the more promising is based on importance sampling. Fundamentally, importance sampling is based on the notion of modifying the underlying probability distribution in such a way that the rare events occur much more frequently. But to use importance sampling, we need to have a deep knowledge of the studied system and, even in such a case, importance sampling may not provide any speed-up. An alternative way to increase the relative number of visits to the rare event is to use trajectory splitting, based on the idea that there exist some well identifiable intermediate system states that are visited much more often than the target states themselves and behave as gateway states to reach the target states [Town98]

The report is organized as follows: Section 2 gives a brief overview and classification of techniques used in Monte Carlo simulations. The next sections provide more detailed description of the most interesting rare event Monte Carlo approaches. Section 3 presents the importance sampling and the sequential importance sampling techniques in Monte Carlo simulation. The importance splitting techniques and the RESTART method are described in section 4. A new method based on interacting particle systems algorithm is introduced in section 5. Finally, section 6 contains conclusions and suggestions for the further work.

2 Monte Carlo simulation techniques

The main techniques that are used by the different methods studied so far within HYBRIDGE are: importance sampling, control variables, multi-level crossing, conditioning, N-particles, stopping time based decomposition, resampling, splitting, and observation filtering. For each of these techniques a short explanation is given below. Moreover, in Table 1 it is shown which of these are used by the different methods studied within HYBRIDGE.

Importance sampling
Assume that we want to estimate $\mathbb{E}[f(x)]$ and suppose the underlying probability density of random variable x is $p(x)$. Then main idea of importance sampling method is to use another probability density $q(x)$. When trying to estimate the mean value for an arbitrary function $f(x)$, we have

$$\mathbb{E}[f(x)] = \int f(x)p(x)dx = \int f(x)\frac{p(x)}{q(x)}q(x)dx.$$

Hence,

$$\mathbb{E}[f(x)] = \mathbb{E}[\frac{f(y)p(y)}{q(y)}]$$

where the random variable y has the probability density $q(y)$. We therefore have the method of estimating $\mathbb{E}[f(x)]$ by using n trials of y, $y_1, \ldots, y_n$ and by approximating $\mathbb{E}[f(x)]$ by

$$\frac{1}{n}\sum_{i=1}^{n}\frac{f(y_i)p(y_i)}{q(y_i)}.$$

Good choice of $q(x)$ can produce an estimate with far lower variance. The main drawback of this method is that it requires a good choice of $q(x)$, and thus a thorough analysis of the problem under consideration.

Control Variables
We write $\mathbb{E}[f(x)]$ in form

$$\mathbb{E}[f(x)] = \mathbb{E}[f(x) - h(x)] + \mathbb{E}[h(x)],$$

where $\mathbb{E}[h(x)]$ can be evaluated analytically and $Var[f(x)-h(x)]$ is appreciably smaller than $Var[f(x)]$. We then use a Monte Carlo method to evaluate $\mathbb{E}[f(x)-h(x)]$.

Conditioning
Suppose that we want to estimate

$$\mathbb{E}[f(x,y)] = \int f(u,v)p_{x,y}(u,v)dudv,$$

where $p_{x,y}(u,v)$ is the density function of the pair (x,y). If we set:

$$h(x) = \frac{1}{m(x)}\int f(x,v)p_{x,y}(x,v)dv = \mathbb{E}[f(x,y)|x],$$

with $m(x) = \int p_{x,y}(x,v)dv$, it is easy to see that $\mathbb{E}[f(x,y)] = \mathbb{E}[h(x)]$. In effect, the distribution of x is $m(x)dx$, and therefore,

$$\mathbb{E}[h(x)] = \int m(u)h(u)du = \int du \int f(x,y)p_{x,y}(u,v)dv = \mathbb{E}[f(x,y)].$$

One can prove that

$$Var(h(x)) \leq Var(f(x,y)).$$

If we can explicitly evaluate the function $h(\cdot)$, it is preferable to use a Monte Carlo simulation for $h(x)$.

Multi-Level crossing
Suppose the target rare set D is contained within a sequence of nested subsets of the state space S, i.e.:

$$D = G_{n+1} \subset \cdots \subset G_1 = S. \tag{1}$$

We denote by $L_i = \partial G_i$ the boundary of the set G_i. If the initial state of process starts in $G_1 \backslash G_2$, then the nesting (1) implies that the rare set can only be reached through crossing all the boundaries (levels) L_i $(i = 2, \ldots, n+1)$.

Stopping time based decomposition
The original problem is decomposed into conditional problem. It is done by introduction of a conditioning on an event and the moment that event happens (the stopping time). For example, suppose that $\{x_t\}$ is a strong Markov process, then for any stopping time τ the following equation holds:

$$\begin{aligned}
\mathbb{E}[f(x_{\tau+t})] &= \int f(x)\mathbb{P}_{x_{\tau+t}}(x \in dx) = \int f(x) p_{x_{\tau+t}}(x)dx \\
&= \int\int f(x) p_{x_{t+\tau}|x_\tau}(x|y) p_{x_\tau}(y)dxdy \\
&= \int \mathbb{E}[f(x_{\tau+t})|x_\tau = y] p_{x_\tau}(y)dy
\end{aligned}$$

here we have assumed that conditional density $p_{x_t|x_\tau}(x|y)$ exists and is uniquely characterized for all x and y.

N-particles
Let x_t denote the state of the system. The N-particles Monte Carlo simulation consists of approximating the density $p_{x_t}(x)$ by a large set of N particles $\left\{x_t^{(i)}\right\}_{i=1}^N$, where each particle has an assigned relative weight, $\omega_t^{(i)}$, such that all weights sum to unity. The density $p_{x_t}(x)$ can be approximated by the empirical distribution:

$$p_{x_t}(x)dx \approx \sum_{i=1}^N \omega_t^{(i)} \delta_{x_t^{(i)}}(dx),$$

where δ denotes the Dirac measure, $\delta_{x_t^{(i)}}(B) = 1$ if $x_t^{(i)} \in B$ and 0 otherwise.

Resampling
During resampling N-particles are independently drawn from the empirical distribution generated by the original set of particles. Resampling is done to reduce the number of particles with almost zero weight, in favour of the particles with significant weight.

Splitting
Entering some intermediate state, which is usually characterized by crossing a threshold (level) by a control parameter, triggers the splitting of the trajectory. The current system state is saved and a number of independent subtrajectories are simulated from the state. Main drawback: difficult to find the optimal splitting parameters (splitting level and number of splits).

Observation filtering
Let $x_t \in \mathbb{R}^n$ denote the state of the observed system and y_t be the observation at time t, then the filtering density $p_{x_t|y_t}(x|y)$ satisfies Bayes theorem:

$$p_{x_t|y_t}(x|y) = \frac{p_{y_t|x_t}(y|x)p_{x_t}(x)}{p_{y_t}(y)}.$$

If the density function $p_{x_t|y_t}(x|y)$ is approximated by the empirical density of set of N particles $\left\{x_t^{(i)}\right\}_{i=1}^{N}$, where each particle has an assigned relative weight, $\omega_t^{(i)}$, such that all weights sum to unity, then Bayes theorem updates the weights as follows:

$$\widehat{\omega}_t^{(i)} = \frac{\omega_t^{(i)} p_{y_t|x_t^{(i)}}(y|x)}{c_t}$$

where c_t is a normalizing constant.

Table 1 Techniques used by Monte Carlo simulation methods so far considered within Hybridge WP2 and WP8

Distinguishing features	Monte Carlo approach							
	A	B	C	D	E.1	E.2	F	G
	§2.1	§2.2	§3.1	§3.2	§4.2	§4.2	WP2.2	[Blom& Bloem03]
Applicability to Hybrid processes	–	Yes	–	–	–	–	Yes	Yes
Importance sampling	Yes	Yes	–	–	–	–	–	–
Splitting	–	–	Yes	Yes		Yes	–	–
Multi-Level crossing	–	–	Yes	Yes	Yes	Yes	–	–
Stopping time based decomposition	–	–	–	–	Yes*	Yes*	Yes*	–
N particles	–	Yes	–	–	Yes	Yes	–	Yes
Control variables	–	–	–	–	–	–	–	–
Resampling	–	Yes	–	–	Yes	–	–	Yes
Observation filtering	–	–	–	–	Yes	Yes	–	Yes
Objective	Rare event	Rare event	Rare event	Rare event	Rare event	Rare event	Rare event	Bayesian filtering

*) applicable only for strong Markov processes

3 Importance Sampling

The Monte Carlo method of simulation is based on the famous law of large numbers which states that given a family of independent random variables $(X_i, i \geq 1)$ with the same distribution as another random variable X such that $\mathbb{E}(|X|) < \infty$ is finished, then with probability one

$$\mathbb{E}(X) = \lim_{N \to \infty} \frac{1}{N}(X_1 + \cdots + X_N).$$

So for large N, we can approximate the expectation of X by the estimate

$$m_N = \frac{1}{N}(X_1 + \cdots + X_N).$$

Nevertheless, the law of large number says nothing about the question: How large N needs to be taken in order to make sure that the absolute value of the error

$$\epsilon_N = \mathbb{E}(X) - m_N$$

is smaller than some given $\epsilon > 0$?

The Central Limit Theorem (CLT) gives an asymptotic random value of error ϵ_N, when $\mathbb{E}(X^2) < \infty$. Let σ^2 the variance of the random variable X, then

$$\frac{\sqrt{N}}{\sigma}\epsilon_N$$

converges in law to centered Gaussian variable with variance one, also called a standard Gaussian variable. Let notice that if G is a standard Gaussian variable, then

$$\mathbb{P}(|G| \leq 1.96) \approx 0.95.$$

We deduce a 95%-confidence interval of the following type

$$[m_N - 1.96\frac{\sigma}{\sqrt{N}},\ m_N + 1.96\frac{\sigma}{\sqrt{N}}].$$

Let consider a simple example of estimation of the probability $p = \mathbb{P}(A)$ of some event A. We have $p = \mathbb{E}(X)$, where X is a Bernoulli random variable taken the value 1 if $X \in A$ and zero otherwise. So, to obtain an approximation of p, we simulate N trials and count the number N_A of occurrences of the event A, in such a way that $p_N = N_A/N$ gives an estimate of p. At the same time, the CLT gives the following confident interval of level 0.95

$$|p - p_N| \leq 1.96\frac{\sigma}{\sqrt{N}} = 1.96\frac{\sqrt{p(1-p)}}{\sqrt{N}} \leq 1.96\frac{1}{2\sqrt{N}} \approx \frac{1}{\sqrt{N}}.$$

So, if A is a rare event $\sigma^2 \approx p$ and the relative error

$$\frac{|p - p_N|}{p} \approx \frac{2}{\sqrt{pN}}.$$

For example, we see that if p is of order 10^{-9}, it is nearly impossible to obtain a reliable statistical estimation. A common approach to speed-up the simulation is to use *importance sampling* techniques.

The basic notion behind importance sampling can be illustrated using a simple example. Let assume that we want to estimate $\mathbb{E}(g(X))$ where X is a random variable with a probability density $f(x)$. Then, the quantity to estimate is given by

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x)f(x)dx.$$

Now, let consider another probability density $h(x)$, such that $h(x) > 0$. Clearly, $\mathbb{E}(g(X))$ can be written

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} \frac{g(x)f(x)}{h(x)}h(x)d.$$

This means that

$$\mathbb{E}(g(X)) = \mathbb{E}\big[\frac{g(Y)f(Y)}{h(Y)}\big],$$

where the random variable Y has the probability density $h(x)$. The goal of this change of probability is to produce an estimate with lower variance. In fact, choosing $h(x) = g(x)f(x)/(\int f(x)g(x)dx)$ yields a zero variance estimator. In practice, however, this is not a feasible change of measure since it requires prior knowledge of $\mathbb{E}(g(X))$. Basically, good variance reduction is achieved by making the ratio $f(x)/h(x)$ very small on the rare set. Heuristically, a good choice can consist in choosing $h(x)$ close to $|g(x)f(x)|$ and normalizing the result. To avoid computing the normalizing constant, we adopt the following estimate

$$\hat{g}_N := \frac{\sum_{i=1}^N g(y_i)f(y_i)/h(y_i)}{\sum_{i=1}^N f(y_i)/h(y_i)} = \sum_{i=1}^N g(y_i)\omega_i, \tag{2}$$

where $y_1, \cdots, y_N$ are i.i.d. random variables with common density h and the *importance weights* are given by

$$\omega_i = \frac{f(y_i)/h(y_i)}{\sum_{i=1}^N f(y_i)/h(y_i)}.$$

For N finite, this estimator is biased but asymptotically, under weak assumptions, the law of large numbers applies, that is $\hat{g}_N$ converges to $\mathbb{E}(g(X))$ as $N \to \infty$. Under additional assumptions, a CLT can be also obtained.

Let remark, that $\hat{g}_N$ in (2) is nothing but the function $g(x)$ integrated with respect to the empirical measure, where δ_a stands for the Dirac measure at a

$$\sum_{i=1}^N \omega_i \delta_{y_i}(dy).$$

Importance sampling is a general Monte Carlo integration method. However, it is not adequate for recursive estimation, mainly when we consider a random trajectory $x_{0;N} = (x_0, \cdots, x_N)$ instead of a real value. Avoiding the computation of the importance weights each time a new data x_{N+1} become available, is the goal of the *sequential importance sampling method.*

3.1 Importance sampling for diffusions

We consider a n-dimensional diffusion $(X_t, 0 \leq t \leq T)$ solution of the stochastic differential equation

$$dX_t = b(t, X_t)dt + \sigma(t, X_t).dB_t$$

where $b(t,x) \in \mathbb{R}^n$, $\sigma(t,x) \in \mathbb{R}^{n\times n}$ and B_t a n-dimensional Brownian motion for the probability $\mathbb{P}$. We assume that the drift vector b and the dispersion matrix σ satisfy the right requirements assuming the existence and uniqueness of solutions of the stochastic differential equation [Kara00].

Given a real function f with polynomial growth, we define the following function

$$u(t,x) = \mathbb{E}(f(X_T)|X_t = x). \quad T \geq t$$

A Monte Carlo simulation consists of approximating $u(T,x)$ by

$$u(t,x) \approx \frac{1}{N}\sum_{k=1}^{N} f(X_T^{(k)}) \tag{3}$$

where $(X_T^{(k)}, k = 1, \cdots, N)$ are independent realizations of the process X at time T, where $X_t^{(k)} = x$. These realizations can be obtained by using the Euler scheme

$$\hat{X}_{t+(i+1)h} = \hat{X}_{t+ih} + b(t+ih, \hat{X}_{t+ih})h + \sigma(t+ih, \hat{X}_{t+ih})(B_{(i+1)h} - B_{ih}).$$

Let notice that the random variables $(B_{(i+1)h} - B_{ih}, k \geq 0)$ are independent centered Gaussian random variables with covariance matrix hId (Id being the identity matrix of $\mathbb{R}^n$).

Given a square integrable $\mathbb{R}^n$-valued process of the form $h(t, X_t)$, adapted to the Brownian motion B_t, we consider the following process Q_t

$$Q_t = \exp\{\int_0^t h(s, X_s).dB_s + \frac{1}{2}\int_0^t \|h(s, X_s)\|^2 ds\}.$$

If $\mathbb{E}(Q_t^{-1}) = 1$, then a new probability measure $\tilde{\mathbb{P}}$ can be defined by the density:

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = (Q_T)^{-1}.$$

With respect to this new measure, the function $u(t,x)$ can be written

$$u(t,x) = \tilde{\mathbb{E}}(f(X_T)Q_T|X_t = x) \tag{4}$$

By Girsanov's theorem the process

$$\tilde{B}_t = B_t + \int_0^t h(s, X_s)ds$$

is a standard n-dimensional Brownian motion under the probability measure $\tilde{\mathbb{P}}$, adapted to the filtration of the Brownian motion B. In terms of this new Brownian motion the processes X and Q can be rewritten as:

$$\begin{aligned} dX_t &= (b(t, X_t) - \sigma(t, X_t)h(t, X_t))dt + \sigma(t, X_t).d\tilde{B}_t \\ dQ_t &= Q_t h(t, X_t).d\tilde{B}_t \end{aligned}$$

which will be used in the simulations for the approximation of (4) by

$$u(t, x) \approx \frac{1}{N}\sum_{k=1}^{N} f(X_T^{(k)})Q_T^{(k)}. \tag{5}$$

We must now determine the functions $h(t, x)$ which lead to a smaller variance for the Monte Carlo approximation given in (5) than for the variance of (3) Firstly, let mention that the function $u(t, x)$ satisfies the Cauchy problem [Kara00][Th 5.7.6]

$$\begin{cases} -\frac{\partial u}{\partial t} & = L_t u; \quad (t, x) \in [0, T) \times \mathbb{R}^n \\ u(T, x) & = f(x); \quad x \in \mathbb{R}^n \end{cases} \tag{6}$$

where L_t is the second-order differential operator

$$L_t u(x) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}(t, x)\frac{\partial^2 u(x)}{\partial x_i \partial x_k} + \sum_{i=1}^{n} b_i(t, x)\frac{\partial u(x)}{\partial x_i},$$

where $a_{ij}(t, x)$ are the components of the diffusion matrix, defined by $a = \sigma\sigma^T$ (σ^T denotes the transpose of σ). Now, using Itô formula, we obtain that

$$\begin{aligned} d(u(t, X_t)Q_t) &= u(t, X_t)Q_t h(t, X_t) + Q_t\sigma^T(t, X_t)\nabla u(t, X_t).d\tilde{B}_t \\ &= Q_t(\sigma^T(t, X_t)\nabla u(t, X_t) + uh)(t, X_t).d\tilde{B}_t \end{aligned}$$

where ∇u denotes the gradient of u with respect to the variable s. Integrating between 0 and T, we obtain

$$u(T, X_T)Q_T = f(X_T)Q_T = u(0, x) + \int_0^T Q_t(\sigma^T(t, X_t)\nabla u(t, X_t) + uh)(t, X_t).d\tilde{B}_t.$$

Therefore the variances in the two Monte Carlo simulations are given by

$$\begin{aligned} \mathrm{Var}_{\tilde{\mathbb{P}}}(f(X_T)Q_T) &= \tilde{\mathbb{E}}\{\int_0^T Q_t^2 \|\sigma^T\nabla u + uh\|^2 dt\} \\ \mathrm{Var}_{\mathbb{P}}(f(X_T)) &= \mathbb{E}\{\int_0^T \|\sigma^T\nabla u\|^2 dt\}. \end{aligned}$$

If $u(t, X_t)$ was known, the problem could be solved and the optimal choice for h, which gives a zero variance, would be:

$$h = -\frac{1}{u}\sigma^T \nabla u.$$

In the practice, we use an approximation for the unknown u in the previous formula which gives a function h such that the Girsanov's theorem applies and the variance of Q_t can be controlled.

The reader could refer to [Four97] where a such approximate is obtained by large deviations techniques, and to [Fouq01] for some examples.

3.2 Sequential Importance Sampling

Let consider a random trajectory $x_{0;N} = (x_0, \cdots, x_N)$ where each x_i is an element of a some state space of the studied system. We want to compute the expectation $\mathbb{E}(h(x_{0;N}))$ with respect to the density $\pi_N(x_{0;N})$. The importance sampling method gives an estimate by generating the trajectories $x_{0;N}^{(i)}$ with respect a measure $q(x_{0;N})$ and computing

$$\hat{h}_m = \frac{\sum_{i=1}^m \omega_N^{(i)} h(x_{0;N}^{(i)})}{\sum_{i=1}^m \omega_N^{(i)}}, \qquad \omega_N^{(i)} = \frac{\pi(x_{0;N}^{(i)})}{q(x_{0;N}^{(i)})}\ . \tag{7}$$

Here, each trial modifies the entire trajectory. However, we can imagine a sequential sampling consisting in updating the trajectory without modifying the past, i.e. $(x_{0;n+1}^{(i)}) = (x_{0;n}^{(i)}, x_{n+1}^{(i)})$. Then, we use the the following identity

$$\pi_N(x_{0;N}) = \pi_0(x_0)\pi_1(x_1|x_0)\cdots\pi_N(x_N|x_{0;N-1}),$$

so the algorithm does the following: Pick a $X_0 = x_0$ according the π_0-density and let $X_{n+1} = x_{n+1}$ according to the conditional π_{n+1}-density of the value x_{n+1} given $x_{0;n}$. In the practice, we change the probability and proceed by sampling X_0 according to a known q_0 density close of π_0 and recursively, x_n is sampled according to the known conditional density $q_n(x_n|x_{0;n-1})$, in such a way that the density of the trajectory $x_{0;N}$ is given by

$$q_N(x_{0;N}) = q_0(x_0)q_1(x_1|x_0)\cdots q_N(x_N|x_{0;N-1}).$$

The importance weights $\omega_N^{(i)} = \pi_N(x_{0;N}^{(i)})/q_N(x_{0;N}^{(i)})$ need to be evaluated, nevertheless this evaluation can be done recursively in time, since

$$\omega_{n+1}^{(i)} = \omega_n^{(i)} \frac{\pi_{n+1}(x_{n+1}^{(i)}|x_{0;n})}{q_{n+1}(x_{n+1}^{(i)}|x_{0;n})}$$

Let remark that the importance weights can only be known up to a proportionality constant.

The problem encountered in the sequential importance sampling method is that, as n increases, the distribution of the importance weights $\omega_n^{(i)}$ becomes more and more skewed. Practically, after a few time steps, only one trajectory has a non-zero importance weight. To avoid this degeneracy, one needs to introduce and additional selection step [Douc01]. The key idea is to eliminate trajectories having low importance weights and to multiply trajectories having high importance weights. Formally, we replace the weighted estimator (7) by the unweighted estimator

$$h_m = \frac{1}{m} \sum_{i=1}^{m} m_N^{(i)} h(x_{0;N}^{(i)}),$$

where $m_N^{(i)}$ is the number of offspring associated to trajectory $x_{0;N}^{(i)}$; it is an integer number such that $\sum_{i=1}^{m} m_N^{(i)} = m$. The $m_N^{(i)}$ are chosen such that h_m is close to $\hat{h}_m$. There are many different way to select $m_N^{(i)}$, for example by sampling according to a m-multinomial distribution of parameters $\omega_N^{(i)}, i = 1, \cdots, m$. The algorithm description is the following [Douc01]

1. *initialization*, $n = 0$

   - For $i = 1, \cdots, m$, sample $x_0^{(i)} \sim q_0(x_0)$ and set $n = 1$.

2. *Importance sampling set*

   - For $i = 1, \cdots, m$, sample $\tilde{x}_n^{(i)} \sim q_n(x_n | x_{0;n-1}^{(i)})$ and set $\tilde{x}_{0;n}^{(i)} = (x_{0;n-1}^{(i)}, \tilde{x}_n^{(i)})$,
   - For $i = 1, \cdots, m$ evaluate the importance weights

   $$\omega_n^{(i)} = \frac{\pi_n(x_n^{(i)} | x_{0;n-1})}{q_n(x_n^{(i)} | x_{0;n-1})}. \tag{8}$$

   - Normalize the importance weights.

3. *Selection step*

   - Resample with replacement m trajectories $(x_{0;n}^{(i)}; i = 1, \cdots, m)$ from the set $(\tilde{x}_{0;n}^{(i)}; i = 1, \cdots, m)$ according to the importance weights.
   - Set $n \leftarrow n + 1$ and go to step 2.

Note that in equation (8), $\omega_{n-1}^{(i)}$ does not appear because the trajectory $x_{0;n-1}^{(i)}$ have uniform weights after the resampling step at time $n - 1$.

4 Importance Splitting techniques

We present here another group of methods to improve the standard simulation. The fundamental idea of trajectory splitting is based on the assumption that there exist some well identifiable intermediate system states that are visited much more often than the rare target event. In trajectory splitting, the step-by-step evolution of the system follows the original probability measure. Entering the intermediate states, which is usually characterized by crossing a threshold by a control parameter, triggers the splitting of the trajectory. The current system state is saved and a number of independent subtrajectories are simulated from the state [Town98].

For example, let consider $m+1$ sets B_i such that the rare event $A = B_{m+1} \subset \cdots \subset B_1$ and use the formula

$$\mathbb{P}(A) = \mathbb{P}(A|B_m)\mathbb{P}(B_m|B_{m-1}) \cdots \mathbb{P}(B_2|B_1)\mathbb{P}(B_1), \tag{9}$$

where each conditioning event on the right hand side of equation (9) is "not rare". The splitting technique proceeds as follows [Tuff00]: make a Bernoulli trial to see if the set B_1 is hit. If it is the case, we split this trial in R_1 Bernoulli trials and we look for each new trial if B_2 is hit. This procedure is repeated at each level, i.e. we make R_i retrials each time B_i is hit by a previous trial. If a level is not hit, neither is A, so we stop the current retrial. Using R_0 independent replications of this procedure, we have then considered $R_0 R_1 \cdots R_m$ trials, considering for example that if we have failed to reach a level B_i at the i-th step, the $R_i \cdots R_m$ possible retrials have failed. An unbiased estimator of $\mathbb{P}(A)$ is

$$\hat{p} = \frac{1}{R_0 \cdots R_m} \sum_{i_0=1}^{R_0} \cdots \sum_{i_m=1}^{R_m} \mathbf{1}_{i_0} \mathbf{1}_{i_0 i_1} \cdots \mathbf{1}_{i_0 i_1 \cdots i_m} = \frac{N_A}{R_0 \prod_{i=1}^{m} R_i},$$

where $\mathbf{1}_{i_0 i_1 \cdots i_j}$ is the result of the i-th Bernoulli retrial at stage j and N_A the total number of trajectories having reached the set A. It can be proven [Vill97] that the optimal simulation is obtained if

$$m = -1/2 Log(P(A)) - 1, \quad P(B_i|B_{i-1}) = e^{-2}, \quad R_i \approx 1/P(B_i|B_{i-1}) = e^2.$$

Nevertheless, in practice the trajectory splitting method may be difficult to apply. For example, the case of the estimation of the probability of a rare event in dynamical system is more complex, since the difficulty to find theoretically the optimal B_i and R_i for each level i. Furthermore, the probability to reach B_i varies generally with the state of entrance in level B_{i-1}.

4.1 Splitting for Markovian models

Now, we assume that the dynamic of the studied system is Markovian and that the initial state O of the system is recurrent, i.e. almost surely the system returns to O in a finite time, (so the system returns to O infinitely often).

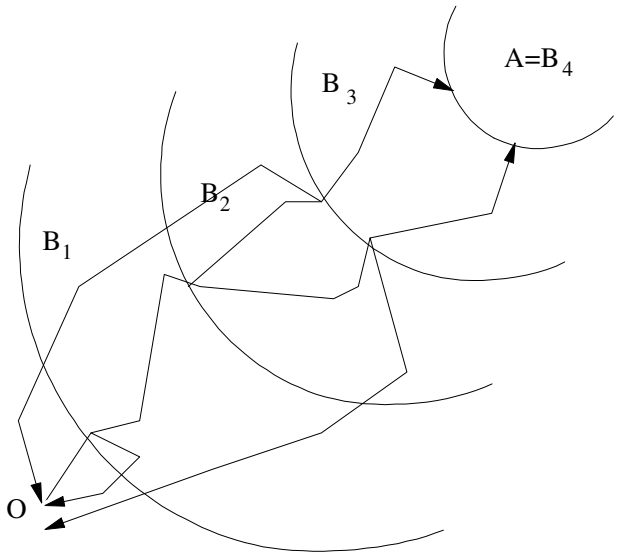

Figure 1: Example with $R_i = 2$ at each level

We want to estimate, using splitting method, the probability p of reaching a rare event A before returning to O. For example, if T_A denotes the first hitting time to A and R the first return time to O, then $p = \mathbb{P}(T_A < R)$ and $\mathbb{E}(T_A) = \mathbb{E}(\min(R, T_A))/p$. So, we proceed as illustrated in Figure 1: if B_1 is hit before going back to O, we split the path in R_1 trials, otherwise, if we are back to O first, we stop the simulation. For each path starting from the point of entrance in B_1, if we hit B_2 before returning to O, then we split this path in R_2 trials, otherwise we stop this path. We do the same thing at each level B_k, and finally, a path from level B_k which hits A before returning to O is considered as a success, so we stop also this path. An implicit assumption is that the level $i + 1$ cannot be reached from level $i - 1$ without entering level i.

In practice [Tuff00], the simulation time is limited to a value T, so that we estimate $\mathbb{P}(T_A < \min(R, T))$. Moreover, going back from B_i to O may take a long time, so we can also gain in simulation time by stopping the simulation of the retrial splitted at level i when it is back d levels down. By then, we assume that it will not hit B_{i+1}, nevertheless this induces a bias which is difficult to estimate.

4.2 RESTART method

We briefly describe in this Section the RESTART method developed by Manuel & José Villén-Altamirano [Vill91], based on the same idea as splitting. RESTART can be used to estimate rare transient events, or the probability $\mathbb{P}(A)$ of every

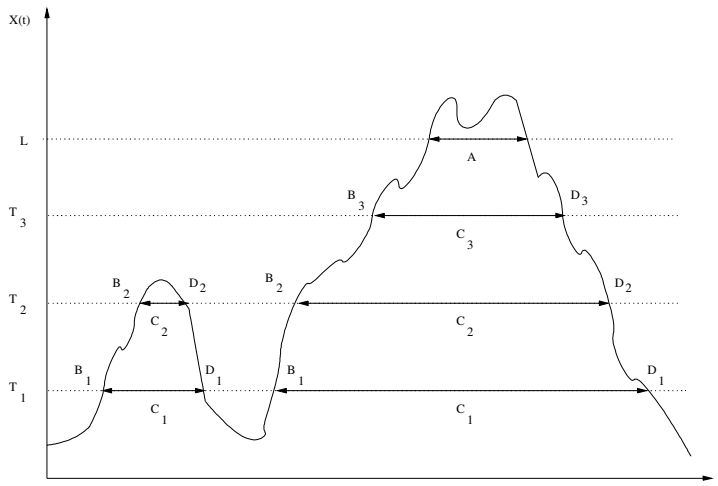

Figure 2: Crude simulation

kind of rare event in steady state, not only the probability of reaching A before coming back to O.

Suppose a given rare event A whose probability is to be estimated. For clarity, we assume that event A can only occur when a system state parameter $X(t)$ up-crosses a threshold L. In Figure 2 an example of system evolution in a crude simulation is given.

Now, let introduce M events C_i (satisfying $A \subset C_M \subset \cdots \subset C_1$) with associated thresholds $T_1, \cdots, T_M$. We define the additional events B_i as the transition from C_{i-1} to C_i (upcrossing of the threshold T_i) and D_i as the transition from C_i to C_{i-1} (downcrossing of the threshold T_i).

The RESTART method illustrated in Figure 3 involves the following procedure [Vill00]:

- A simulation path, called main trial, is performed in the same way as if it were a crude simulation (thick path in Figure 3). It lasts until it reaches a predefined "end of simulation" condition (e.g. a predefined value of the simulated time).

- Each time an event B_1 occurs in the main trial, the system state X_{B_1} is saved, the main trial is interrupted and $R_1 - 1$ retrials at level 1 are performed. Each retrial is a simulation path that starts from the state X_{B_1} and finishes when an event D_1 occurs.

- After the $R_1 - 1$ retrials of level 1 have been performed, the main trial continues from the state X_{B_1}. Note that the total number of simulated paths $[B_1, D_1)$ is R_1 (including the main trial). The main trial which continues after D_1 leads to new sets of retrials of level 1.

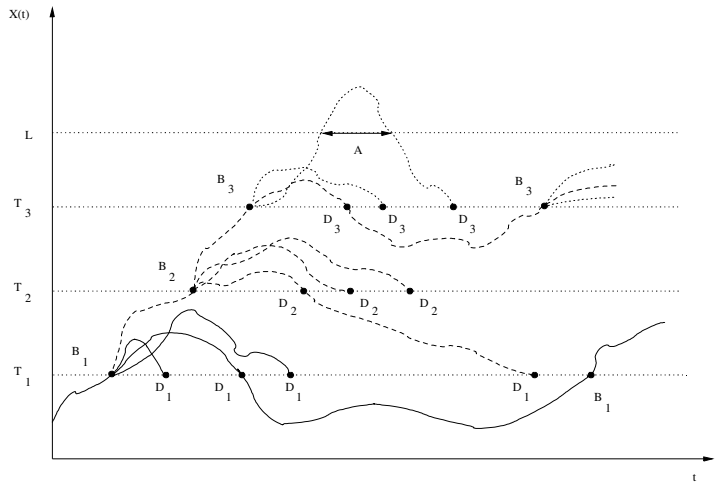

Figure 3: Simulation with RESTART

- If an event B_2 occurs during any trial $[B_1, D_1)$, an analogous process is done: $R_2 - 1$ retrials of level 2 starting in B_2 and finishing in D_2 are performed, leading to R_2 trials $[B_2, D_2)$. The trial $[B_1, D_1)$ which continues after D_2 may lead to new sets of retrial of level 2 if new events B_2 occur.

- In general R_i trials $[B_i, D_i)$ $(1 \leq i \leq M)$ are performed each time an event B_i occurs in a trial $[B_{i-1}, D_{i-1})$. The number R_i is constant for each value i

- A retrial of level i also finishes if it reaches the "end of simulation " condition before the occurence of event D_i.

The differences with respect to the 4.1 splitting lie in the fact that we consider the path as finished if it leaves the level i towards level $i-1$ (it is the case where we diminish the computation time with $d = 1$ in the previous section), except for the last retrial at level i. As a matter of fact, the last is authorized to go under level i and to continue until it reaches again level i and is split again. Nevertheless the weight of the last path is multiplied by R_i if it reaches a lower level before going to an upper one, by then we do as if all the retrials were grouped again together in a single one.

Let N_A be the total number of events A that occurs in the simulation (in the main trial and any retrial) and N be the number of reference events simulated in the main retrial, then the unbiasedness of the estimator of $\mathbb{P}(A)$

$$\hat{P} = \frac{N_A}{N \prod_{i=1}^{M} R_i}$$

is proved in [Vill00]. The method used is based on the following recurrent property: Consider a simulation with M thresholds. If the retrials of level 1 (and their corresponding upper-level retrials) are not taken into account, we obtain a simulation with $M-1$ thresholds (T_2 to T_M).

In [Vill00], optimal values for thresholds and the number of retrials that maximize the gain obtained with RESTART are also derived (the gain is defined as the ratio of the product of the simulation cost by the variance of the estimator for a crude simulation by the same product but computing for a RESTART simulation). Nevertheless, it seems that some assumptions used in this paper should be clarified.

5 Interacting Particle Systems Algorithm

We present now a genetic type interacting particle systems algorithm and a genealogical model for estimating a class of rare events, following [Cérou02]. We consider a continuous-time strong Markov process which is assumed to start in some Borel set O with a given initial probability distribution η_0. For a given target Borel set A, we define the first time the process X hits A, namely

$$T_A = \inf\{t \geq 0 \ : \ X_t \in A\}, \quad T_A = \infty \text{ if the set is empty.}$$

We would like to estimate the quantities

$$\mathbb{P}(T_A \leq T) \quad \text{and} \quad \text{Law}(X_t, 0 \leq t \leq T_A | T_A \leq T), \tag{10}$$

where T is either a deterministic finite time or the almost surely finite entrance time into a recurrent Borel set R if $R \cap O = \emptyset$ or the first return time to O if $R = O$, namely

$$T = \inf\{t > 0 \ : \ X_t \in R\}.$$

5.1 Multi-level Feynman-Kac formulae

As previously, the process X, before visiting R or entering into A, passes through a decreasing sequence of Borel level sets

$$A = B_{m+1} \subset \cdots \subset B_1,$$

with $O \cap B_1 = \emptyset$ and $R \cap B_1 = \emptyset$. To capture the behavior of X between the different levels, we introduce the discrete stochastic sequence $\mathcal{X}_n$ representing the paths of X between the successive levels, these paths having various length (see Figure 4),

$$\mathcal{X}_n = (X_t \, ; T_{n-1} \wedge T \leq t \leq T_n \wedge T), \quad \mathcal{X}_0 = X_0,$$

where $a \wedge b = \min(a, b)$ and each T_n represents the first time X reaches B_n, that is

$$T_n = \inf\{t \geq 0 \ : X_t \in B_n\}, \quad 1 \leq n \leq m+1, \qquad T_0 = 0.$$

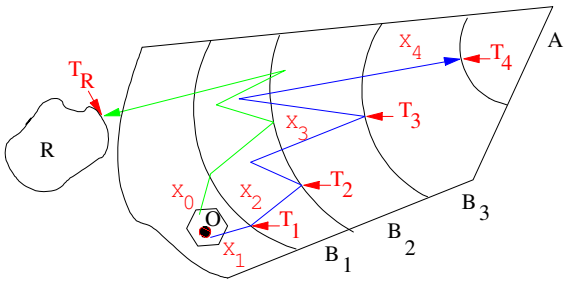

Figure 4: Embedded Markov Chain

If $T < T_{n-1}$, then $\mathcal{X}_n = X_T$ and $X_{T_n \wedge T} = X_T \notin B_n$, whereas if $T_{n-1} \leq T < T_n$, then $\mathcal{X}_n = (X_t, T_{n-1} \leq t \leq T)$ and $X_{T_n \wedge T} = X_T \notin B_n$. Finally, if $T_n \leq T$, then $\mathcal{X}_n = (X_t, T_{n-1} \leq t \leq T_n)$ represents the path of X between the successive levels B_{n-1} and B_n and $X_{T_n \wedge T} = X_{T_n} \in B_n$. We also observe that

$$(T_A \leq T) = (T_{m+1} \leq T) = (T_1 \leq T, \cdots, T_{m+1} \leq T).$$

By the strong Markov property, we can check that the stochastic sequence $(\mathcal{X}_n)$ forms a Markov chain with value in the path-space of X, denoted by E. To check wether the path has succeeded to reach the level B_n, we introduce the functions g_n defined for each path $x_{s;t} := (x_u)_{s \leq u \leq t}$, with $s \leq t$ by

$$g_n(x_{s;t}) = \mathbf{1}_{\{x_t \in B_n\}}.$$

With this notation, we have for each n

$$(T_n \leq T) = (g_1(\mathcal{X}_1) = 1, \cdots, g_n(\mathcal{X}_n) = 1)$$

Then, the following Feynman-Kac representation of the quantities (10) has been proved in [Cérou02]

$$\mathbb{E}(f(\mathcal{X}_n)|T_n \leq T) = \frac{\mathbb{E}(f(\mathcal{X}_n)\prod_{p=0}^n g_p(\mathcal{X}_p))}{\mathbb{E}(\prod_{p=0}^n g_p(\mathcal{X}_p))} \tag{11}$$

$$\mathbb{P}(T_n \leq T) = \mathbb{E}(\prod_{p=0}^n g_p(\mathcal{X}_p)) \tag{12}$$

where f in any bounded function defined on the path-space E of X. In addition, for any $n+1$-valued bounded function f, we have that

$$\mathbb{E}(f(\mathcal{X}_0, \cdots, \mathcal{X}_n)|T_n \leq T) = \frac{\mathbb{E}(f(\mathcal{X}_0, \cdots, \mathcal{X}_n)\prod_{p=0}^n g_p(\mathcal{X}_p))}{\mathbb{E}(\prod_{p=0}^n g_p(\mathcal{X}_p))}.$$

5.2 Genetic approximating models

Based on the Feynman-Kac representation formulae, we can introduce a genetic particle algorithm to approximate the desired quantities (10). These interacting particle systems approximating models for solving a general class of Feynman-Kac models have been studied in [Del Moral 00], [Del Moral 01a] and [Del Moral 01b].

To describe this particle approximating model, let introduce the probability distribution $\eta_n \in \mathcal{P}(E)$ on the space E, defined for each bounded function f on E by

$$\eta_n(f) := \int_E f(x)\eta_n(dx) = \frac{\gamma_n(f)}{\gamma_n(1)}, \text{ with } \gamma_n(f) = \mathbb{E}(f(\mathcal{X}_n)\prod_{p=0}^{n-1} g_p(\mathcal{X}_p)).$$

Thus, the right-hand term of (11) is nothing else that the ratio $\eta_n(fg_n)/\eta_n(g_n)$.

We used this notation, since the evolution of η_n is driven by a non-linear dynamical system

$$\eta_{n+1} = \Phi_{n+1}(\eta_n), \tag{13}$$

where the mappings Φ_{n+1} from the set of measures

$$\mathcal{P}_n(E) = \{\eta \in \mathcal{P}(E); \eta(g_n) > 0\}$$

into $\mathcal{P}(E)$ are defined by

$$\Phi_{n+1}(\eta)(dv) = \int_E \Psi_n(\eta)(du)\mathcal{K}_{n+1}(u, dv),$$

where

$$\Psi_n(\eta)(f) = \frac{\eta(fg_n)}{\eta(g_n)},$$

and $\mathcal{K}_n(u, dv)$ represents the Markov transition kernel of the Markov chain $\mathcal{X}_n$, i.e.

$$\mathcal{K}_n(u, dv) = \mathbb{P}(\mathcal{X}_n \in dv | \mathcal{X}_{n-1} = u).$$

Thus, the recursion (13) involves two separate updating/prediction transitions

$$\eta_n \in \mathcal{P}(E) \xrightarrow{updating} \hat{\eta}_n = \Psi_n(\eta_n) \in \mathcal{P}(E) \xrightarrow{prediction} \eta_{n+1} \in \mathcal{P}(E). \tag{14}$$

Let notice that the measures γ_n on E can be expressed in terms of the sequence $(\eta_p; p \leq n)$, using the easily checked formula

$$\gamma_n(f) = \eta_n(f)\prod_{p=0}^{n-1} \eta_p(g_p).$$

In these notations, we observe that

$$\gamma_n(g_n) = \mathbb{P}(T_n < R),$$

and

$$\hat{\eta}_n(f) = \mathbb{E}(f(\mathcal{X}_n)|T_n \leq T).$$

The genetic type N-particle system associated with an abstract measure valued process of the form (13) is based on the natural idea to approximate η_n for $n \geq 1$ by a sequence of empirical measures

$$\eta_n^N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_n^{(i)}}, \tag{15}$$

associated with a system of N interacting particles $\xi_n = (\xi_n^{(1)}, \cdots, \xi_n^{(N)})$ moving in the set E. In other words, we approximate the two-step transitions of the system (14) by a two-step Markov chain

$$\eta_n^N \xrightarrow{updating} \hat{\eta}_n = \frac{1}{N}\sum_{i=1}^{N} \delta_{\hat{\xi}_n^{(i)}} \xrightarrow{prediction} \eta_{n+1}^N,$$

where the $\hat{\xi}_n^{(i)}$ are N independent random variables with common distribution

$$\Psi_n(\frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_n^{(i)}}) = \sum_{i=1}^{N} \frac{g_n(\xi_n^{(i)})}{\sum_{j=1}^{N} g_n(\xi_n^{(j)})} \delta_{\xi_n^{(i)}}.$$

Using the approximating measures η_n^N, we associate with γ_n the approximating measure γ_n^N defined by

$$\gamma_n^N(f) = \eta_n^N(f) \prod_{p=0}^{n-1} \eta_p^N(g_p). \tag{16}$$

Consequently, an estimate of $\mathbb{P}(T_A \leq T)$ is nothing else that $\gamma_{m+1}^N(g_{m+1})$.

More precisely, the N-particle system associated with the dynamical system (13) is the Markov chain $\xi_n = (\xi_n^{(1)}, \cdots, \xi_n^{(N)})$ taking value in the product spaces $E^N \cup \{\Delta\}$, where Δ stands for a cemetery point. Its transitions are defined as follows. The initial system of particles $\xi_0 = (\xi_0^{(1)}, \cdots, \xi_0^{(N)})$ consists in N independent random variables with common law η_0, the law of X_0. Then, for any configuration $x = (x^{(1)}, \cdots, x^{(N)}) \in E^N$ (a collection of N trajectories), such that the corresponding empirical measure $N^{-1}\sum_{i=1}^{N} \delta_{\xi_n^{(i)}}$ belongs to $\mathcal{P}_n(E)$ the probability that $\xi_{n+1} \in dy^{(1)} \times \cdots \times dy^{(N)}$ given $\xi_n = x$ is the product

$$\prod_{p=1}^{N} \Phi_{n+1}\big(N^{-1}\sum_{i=1}^{N} \delta_{\xi_n^{(i)}}\big)(dy^{(p)}).$$

The rationale behind this is that the N particles $\xi^{(i)}$ evolve independently with the law $\Phi_{n+1}(\eta_n^N)$.

When the system arrives in some configuration $\xi_n = x$ such that $N^{-1}\sum_{i=1}^N \delta_{\xi_n^{(i)}} \notin \mathcal{P}_n(E)$ (i.e. all particles hit R), the algorithm is stopped and we set $\xi_{n+1} = \Delta$.

Furthermore, let introduce I_{n+1}^N the set of the labels of the particles having succeeded to reach the $(n+1)$-th level

$$I_{n+1}^N = \{1 \le i \le N : \xi_{n+1}^{(i)}(T_{n+1}^{(i)}) \in B_{n+1}\}.$$

So I_{n+1}^N is empty if and only if each $N^{-1}\sum_{i=1}^N \delta_{\xi_{n+1}^{(i)}} \notin \mathcal{P}_{n+1}(E)$, and in this situation the algorithm is stopped.

This algorithm can be splitted in two separate transitions; the mutation/prediction transition $\hat{\xi}_{n+1} \to \xi_{n+1}$ and the selection/updating transition $\xi_{n+1} \to \hat{\xi}_{n+1}$ where the $\hat{\xi}_{n+1}^{(i)}$ are N independent random variables with common distribution

$$\Psi_{n+1}(\frac{1}{N}\sum_{i=1}^N \delta_{\xi_{n+1}^{(i)}}) = \sum_{i=1}^N \frac{g_{n+1}(\xi_{n+1}^{(i)})}{\sum_{j=1}^N g_{n+1}(\xi_{n+1}^{(j)})}\delta_{\xi_{n+1}^{(i)}} = \frac{1}{|I_{n+1}^N|}\sum_{i\in I_{n+1}^N} \delta_{\xi_{n+1}^{(i)}}, \quad (17)$$

where $|I_{n+1}^N|$ is the cardinal of the set I_{n+1}^N.

The genetic algorithm is the following:

1. *initialization*, $n = 0$
   - For $i = 1, \cdots, N$, sample $\xi_0^{(i)} \sim \eta_0$ and set $\hat{\xi}_0^{(i)} = \xi_0^{(i)}$.
2. *Mutation/Prediction* $\hat{\xi}_n^{(i)} \to \xi_{n+1}^{(i)}$
   - If $\hat{\xi}_n = \Delta$, then $\xi_{n+1} = \Delta$ (i.e. the algorithm is stopped),
   - otherwise, for $i = 1, \cdots, N$, generate independently a new excursion from $X_{T_n^{(i)}}$ to the $n+1$-th level or the set R, according to the Markov transition of the Markov chain $\mathcal{X}_{n+1}$ at time $n+1$,
   - for $i = 1, \cdots, N$, evaluate the number of excursions having succeeded to reach the $n+1$-th level.
3. *Selection/Updating* $\xi_{n+1}^{(i)} \to \hat{\xi}_{n+1}^{(i)}$
   - Resample the N particles $\xi_{n+1}^{(i)}$ according to the empirical measure (17),
   - set $n \leftarrow n+1$ and go to step 2.

During the mutation/prediction, each particle evolves randomly according to a given transition probability kernel. During the selection/updating transition the particles are selected for reproduction, the most fit individuals being more likely to be selected. Thus this transition allows heavy particles to give birth to some particles at the expense of light particles, which die. To avoid to

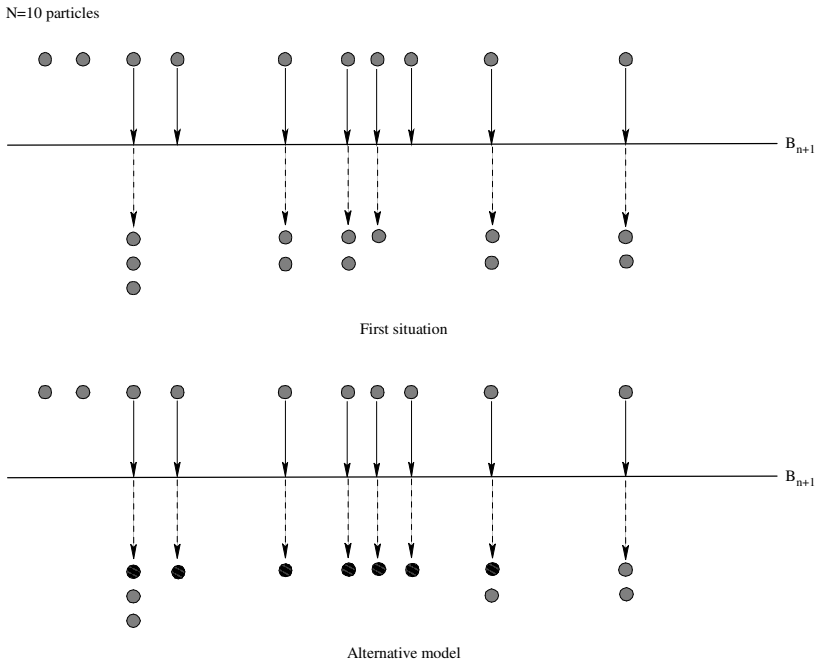

Figure 5: Differences between the two selection/updating models

give greater place to the heavy particles, an alternative scheme can be proposed [Cérou02] by modifying the Selection/Updating transition. Instead of resampling with the empirical measure (17), we use another selection distribution which consists in the following. At step $n+1$, if the particle has succeeded to reach the $n+1$-th level, we keep it, so $\hat{\xi}_{n+1}^{(i)} = \xi_{n+1}^{(i)}$. In the opposite, the terminal point $X_{T_{n+1}^{(i)}} \notin B_{n+1}$ of the path/particle $\xi_{n+1}^{(i)}$ does not belong to B_{n+1}. In this case, $\hat{\xi}_{n+1}^{(i)}$ is chosen randomly and uniformly in the set of the particle having succeeded to reach the level B_{n+1}, that mean the particles whose the labels are in the set I_{n+1}^N. In other world each particle which does not enter into the $(n+1)$-th level is killed and instantly a different particle in the B_{n+1} level splits into two offsprings (see Figure 5 for the differences between the two models)

Let denote by τ^N the life-time of the N-genetic model

$$\tau^N = \inf\{n \geq 0 : \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_n^{(i)}} \notin \mathcal{P}_n(E)\}.$$

For each time $n < \tau^N$, we consider the approximating measure γ_n^N (16), and we

obtain

$$\gamma_n^N(g_n) = \prod_{p=1}^{n} \frac{|I_p^N|}{N}$$

and

$$\hat{\eta}_n^N = \frac{1}{|I_n^N|} \sum_{i \in I_n^N} \delta_{\xi_n^{(i)}}.$$

The particle estimates $\gamma_n^N(g_n)$ are unbiased

$$\mathbb{E}(\gamma_n^N(g_n) 1_{\{\tau^N > n\}}) = \mathbb{P}(T_n \leq T).$$

The asymptotic behavior as $N \to \infty$ of the interacting particle model is described in [Cérou02].

6 Concluding Remarks

Many techniques for estimating rare event probabilities have been reviewed on their applicability to ATM. Two main streams of development have been identified: approximate numerical solution of the integro-partial differential equations and Monte Carlo simulation based approaches. The approximate numerical approaches have been studied in details in task 1 of the work package WP8 [Krys03a]. The most promising rare event Monte Carlo approaches such as importance sampling, importance splitting and interacting particle system methods have been considered in the present report. Importance sampling and importance splitting methods are known to be of great use for the practice. However, these two methods may not be satisfactory when applied to high dimensional complex dynamic problems such as in ATM.

The importance sampling techniques, based on changing probability distributions to make rare events less rare, has been used to obtain dramatic improvements in efficiency in estimating small probabilities in queueing and reliability systems. But the effectiveness of such methods depends critically on the ability to find the right change of measure. If it is done improperly, the importance sampling may produce worse results than straightforward simulation. Finding the right change of measure generally requires identifying at least the rough asymptotics of a rare event probability. This type of analysis can be formidable in complex dynamic models.

The importance splitting methods are quite powerful if one can find the optimal splitting parameters: number and position of thresholds (levels), and the number of splitting's at each level. This can be difficult when the scale of complexity increases.

The interacting particle system algorithm is a new type of splitting technique. It combines the multilevel splitting techniques with the branching and interacting particle systems approximations of Feynman-Kac distributions. There are many interesting questions left, such as the optimal selection of the splitting levels, and how the number of levels, the number of particles and the number of independent simulations influence the accuracy. For accident risk assessment the level of accuracy should be predictable.

Further studies will concentrate on making the accuracy predictable, and on extending the rare event Monte Carlo approaches also to hybrid state Markov process situations. One of the starting points for the latter is the hybrid state particle filtering method that has successfully been developed within WP8.2 for multi-target tracking applications [Blom&Bloem03]. A complementary issue to be studied is to combine complementary methods, e.g. importance sampling and sequential level crossing based splitting, in an effective way.

Acknowledgement: *We would like to thank Fabien Campillo for suggesting a number of improvements.*

References

[Aldo89] Aldous, D. *Probability Approximations via the Poisson Clumping Heuristic*, volume 77 of *Applied Mathematical Sciences*. Springer Verlag, 1989.

[Blom et al 03a] Blom, H.A.P., Bakker, G.J., Everdij M.H.C., M.N.J. van der Park, Stochastic analysis background of accident risk assessment for Air Traffic Management, Report Hybridge WP2.2, January 21st 2003.

[Blom&Bloem03] H.A.P. Blom, E.A. Bloem, Proceedings 6th Int. Conf on Information Fusion, Cairns, Queensland, Australia, 8-11 July 2003, pp. 785-792.

[Cérou02] Cérou, F. & Del Moral, P. & LeGland, F. & Lezaud, P. *Genetic Genealogical Models in Rare Events Analysis*. Preprint, 2002.

[Del Moral 00] Del Moral, P. and Miclo, L. *Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering*. Séminaire de Probabilités XXXIV. Lecture Notes in Mathematics No. 1729, J. Azma, M Émery, M. Ledoux, M. Yor (Eds.), pp. 1-145, 2000.

[Del Moral 01a] Del Moral, P. and Miclo, L. *Genealogies and increasing propagations of chaos for Feynman-Kac and genetic models*. Annals of Applied Probability, vol. 11, No. 4, pp. 1166-1198, 2001.

[Del Moral 01b] Del Moral, P. and Jacod, J. *Interacting Particle Filtering With Discrete Observations* Sequential Monte Carlo Methods in Practice. Springer Verlag, Doucet, A. , de Freitas, N. , Gordon, N. (Eds), 2001.

[Douc01] Doucet, A. and de Freitas, N. and Gordon, N. *An Introduction to Sequential Monte Carlo Methods.* Sequential Monte Carlo Methods in Practice. Springer Verlag, Doucet, A. , de Freitas, N. , Gordon, N. (Eds), 2001.

[Glass99] Glasserman P., Heidelberger P., Shahabuddin P., Zajic T., Multilevel splitting for estimating rare event probabilities, Operations Research, vol.47, issue 4 (Jul.-Aug., 1999), pp. 585-600.

[Fouq01] Fouque, J. P. and Tullie, T. A. *Variance Reduction for Monte Carlo Simulation in a Stochastic Volatility Environment* AMS SMF Congress, Special Session on Mathematical Methods in Financial Modeling, Avellaneda, M. , Cont, R. (Eds), 2001.

[Four97] Fournie, E. and Lebuchoux, J. and Touzi,N. *Small Noise Expansion and Importance Sampling* Asymptotics Analysis, 14(4), 1997, pp. 361-376.

[Kara00] Karatzas, I. and Shreve, S. *Brownian Motion and Stochastic Calculus*. Springer, 2000.

[Krys03a] J. Krystul, A. Bagchi, H. Blom, Risk decomposition and assessment methods, Report Hybridge WP8.1, May 2003.

[Krys03b] J. Krystul, Weak approximation of first exit time, draft work paper, 2003.

[Lape03] Lapeyre B., Pardoux É. and Sentis R., Introduction to Monte-Carlo Methods for Transport and Diffusion Equations, Oxford University Press, 2003.

[Reiss97] Reiss, R. D. & Thomas, M. *Statistical Analysis of Extreme Values*. Birkhäuser, 1997.

[Town98] Townsend, J. D. & Haraszti, Z. & Freebersyser, J. A. & Devetsikiotis, M. *Simulation of Rare Events in Communications Networks*. in *IEEE Communications Magazine, August 1998*

[Tuff00] Tuffin, B. & Trivedi, K. S. *Implementation of Importance Splitting Techniques in Stochastic Petri Net Package*. Computer Performance Evaluation Modelling Techniques and Tools. Springer Verlag, Lecture Notes in Computer Science, No 1786, B.R. Haverkort, H.C. Bohnenkamp, C.U. Smith, pp 216-229, 2000.

[Vill91] Villén-Altamirano, M. and Villén-Altamirano, J. *RESTART: A method for accelerating rare event simulations.* In *13th Int. Teletraffic Congress,ITC 13 (Queueing, Performance and Control in ATM)*, Copenhagen, Denmark, 1991, pp. 71–76.

[Vill97] Villén-Altamirano, M. and Villén-Altamirano, J. *RESTART: An Efficient and General Method for Fast Simulation of Rare Event.* Technical Report 7, Departamento de Mætmatica Aplicada, E.U. Inform'atica, Universidad Politécnica de Madrid, 1997.

[Vill00] Villén-Altamirano, M. and Villén-Altamirano, J. (2000). *Analysis of RESTART Simulation: Theoretical Basis and Sensitivity Study* In *3th Workshop on Rare Event Simulation, RESIM 2000*, October 5-6, Pisa, Italy.