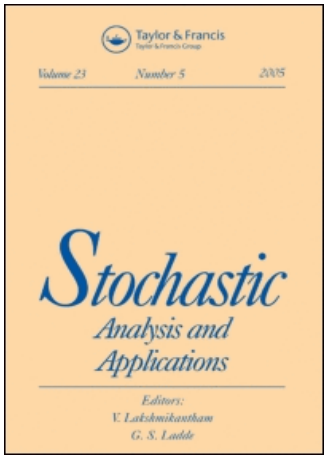
This article was downloaded by:[Cérou, Frédéric]
[Cérou, Frédéric]
On: 6 April 2007
Access Details: [subscription number 772729327]
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Stochastic Analysis and Applications
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597300

Adaptive Multilevel Splitting for Rare Event Analysis

To cite this Article: , 'Adaptive Multilevel Splitting for Rare Event Analysis', Stochastic Analysis and Applications, 25:2, 417 - 443
To link to this article: DOI: 10.1080/07362990601139628
URL: http://dx.doi.org/10.1080/07362990601139628

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

© Taylor and Francis 2007

Stochastic Analysis and Applications, 25: 417–443, 2007
Copyright © Taylor & Francis Group, LLC
ISSN 0736-2994 print/1532-9356 online
DOI: 10.1080/07362990601139628

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

Adaptive Multilevel Splitting for Rare Event Analysis

Frédéric Cérou
IRISA/INRIA, Campus de Beaulieu, Rennes, France

Arnaud Guyader
Equipe de Statistique, Université Haute Bretagne, Rennes, France

Abstract: The estimation of rare event probability is a crucial issue in areas such as reliability, telecommunications, aircraft management. In complex systems, analytical study is out of question and one has to use Monte Carlo methods. When rare is really rare, which means a probability less than 10^{-9}, naive Monte Carlo becomes unreasonable. A widespread technique consists in multilevel splitting, but this method requires enough knowledge about the system to decide where to put the levels at hand. This, unfortunately, is not always possible. In this article, we propose an adaptive algorithm to cope with this problem: The estimation is asymptotically consistent, costs just a little bit more than classical multilevel splitting, and has the same efficiency in terms of asymptotic variance. In the one-dimensional case, we rigorously prove the a.s. convergence and the asymptotic normality of our estimator, with the same variance as with other algorithms that use fixed crossing levels. In our proofs we mainly use tools from the theory of empirical processes, which seems to be quite new in the field of rare events.

Keywords: Multilevel splitting; Quantiles; Rare events.

Mathematics Subject Classification: 65C05; 65C35; 60F05; 62G30.

Received April 6, 2006; Accepted September 22, 2006

This work was partially supported by Inria ARC (Action de recherche coopérative) RARE. The authors would like to thank Philippe Berthet for valuable discussions about empirical processes, and Hélène Topart for her participation to the numerical simulations.

Address correspondence to Frédéric Cérou, IRISA/INRIA, Campus de Beaulieu, Rennes Cédex 35042, France; E-mail: frederic.cerou@inria.fr

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

1. INTRODUCTION

Let $(X_t)_{t\geq 0}$ be a strong Markov process with values in $\mathbb{R}$. Suppose that $X_0 = x_0 > 0$, the origin 0 is an attractive point and $M \gg 0$ is a barrier that the process is very unlikely to reach. We would like to estimate the probability that (X_t) reaches M before coming back to 0: This is our rare event. A typical example is given in Figure 1.

For the sake of simplicity, we suppose that the trajectories are continuous. The point 0 is "attractive" means that if we define

$$T_0 = \inf\{t \geq 0 : X_t = 0\},$$

the stopping time T_0 is such that: $\mathbb{E}[T_0] < +\infty$.

Since (X_t) tends to decrease to 0, it is clear that if we denote

$$T_M = \inf\{t \geq 0 : X_t = M\},$$

we have

$$\mathbb{P}(T_M < T_0) \approx 0.$$

Let us consider the situation when this very small probability $\mathbb{P}(T_M < T_0) = \alpha$ is strictly positive. We want to get an estimation $\hat{\alpha}$ of α.

When the process $(X_t)_{t\geq 0}$ is not simple, the usual way to cope with this kind of problem is to use Monte Carlo techniques. The most natural one is to simulate n i.i.d. trajectories of the process, to count those who reach M before 0, and to compute the ratio. Unfortunately, when the event is really rare, this is completely unrealistic. Another idea, called Importance Sampling, is to simulate trajectories with respect to another reference probability, so that the rare event becomes less rare, and then to correct the estimation via the importance function (which corresponds to a Radon–Nikodym derivative). But it is a difficult problem to find another suitable reference probability, especially when the system in study is very complex.

In such a situation, a classical method is the one of multilevel splitting: the idea dates back to 1951 with the work of Kahn and Harris in the setting of particle transmission [12]. Roughly speaking, the principle is to multiply the trajectories that approach the event of interest and to let the others die. It is much more simple to implement than importance sampling, but it requires some a priori knowledge on the system also; precisely, when and how many times shall one split the trajectories? Hereafter we propose a new method, called adaptive multilevel splitting, which uses the splitting idea, but where the splitting levels are determined during the simulations.

The article is organized as follows. Section 2 describes the algorithm and the estimator of the rare event probability. Section 3 proves

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

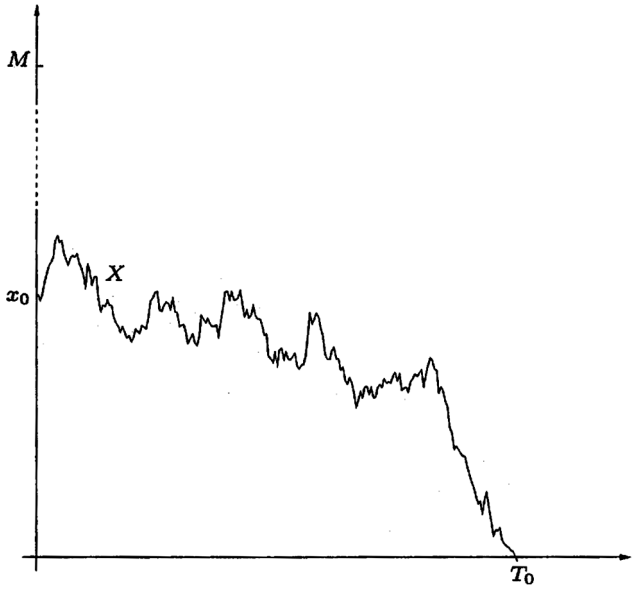


Figure 1. An example of trajectory for the Markov process.

the consistency of the estimator. Section 4 establishes the asymptotic normality, with a very simple expression of the variance. Section 5 illustrates these results on a toy example. Finally, Section 6 compares this adaptive algorithm with existing versions of multilevel splitting and shows how the method can be used in various situations.

2. THE ALGORITHM

The structure of the algorithm is the following: to approach the barrier M of interest, we proceed in several steps. Except the first and the last ones, all steps are equivalent. In what follows, n denotes the number of particles which are simulated, and k denotes the number of particles that we do not throw away from one step to another.

- *Step* 1: simulate n i.i.d. trajectories $\left(X_t^j\right)_{t\geq 0}$ according to the law of the process $(X_t)_{t\geq 0}$ and with common initial condition

$$\forall j \in \{1, \ldots, n\} \quad X_0^j = x_0 > 0.$$

  Wait until all trajectories have reached 0: for the jth particle, this requires time T_0^j, with $\mathbb{E}[T_0^j] = \mathbb{E}_{x_0}[T_0] < +\infty$. Denote

$$S_{n,j}^1 = \sup_{0\leq t\leq T_0^j} X_t^j,$$

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

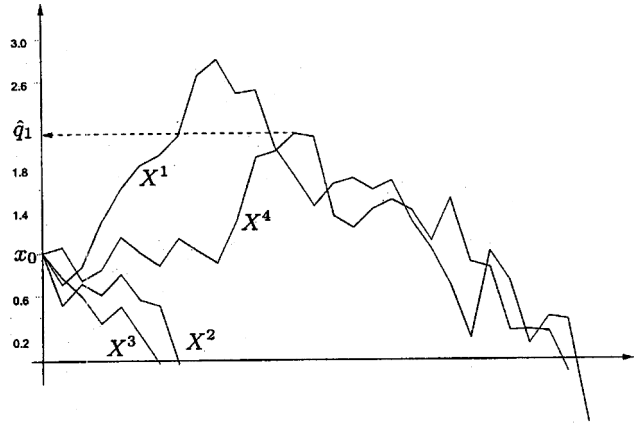


Figure 2. The first step of the algorithm, with $n = 4$ and $k = 1$.

and sort the sample $\left(S^1_{n,1}, \ldots, S^1_{n,n}\right)$ in increasing order:

$$S^1_{n,(1)} \leq \cdots \leq S^1_{n,(n-k)} \leq \cdots \leq S^1_{n,(n)}.$$

Keep in memory the quantity (see Figure 2)

$$\hat{q}_1 = S^1_{n,(n-k)}.$$

- *Step* 2: Keep $\left(S^1_{n,(n-k+1)}, \ldots, S^1_{n,(n)}\right)$ unchanged, but denote them simply $\left(S^2_{n,n-k+1}, \ldots, S^2_{n,n}\right)$. Simulate $(n-k)$ trajectories $\left(X^j_t\right)_{t\geq 0}$ from initial point $\hat{q}_1$. Wait until all these $(n-k)$ trajectories have reached 0: for the jth particle, this requires time T^j_0, with $\mathbb{E}[T^j_0] = \mathbb{E}_{\hat{q}_1}[T_0] < +\infty$. For each $j \in \{1, \ldots, n-k\}$, denote

$$S^2_{n,j} = \sup_{0\leq t\leq T^j_0} X^j_t,$$

and sort the sample $\left(S^2_{n,1}, \ldots, S^2_{n,n-k}, S^2_{n,n-k+1}, \ldots, S^2_{n,n}\right)$ in increasing order:

$$S^2_{n,(1)} \leq \cdots \leq S^2_{n,(n-k)} \leq \cdots \leq S^2_{n,(n)}.$$

Keep in memory the quantity (see Figure 3)

$$\hat{q}_2 = S^2_{n,(n-k)}.$$

- Repeat the procedure until iteration $\widehat{N}$ such that $\hat{q}_{\widehat{N}+1} \geq M$. Among the sample $\left(S^{\widehat{N}}_{n,1}, \ldots, S^{\widehat{N}}_{n,n}\right)$, there is a proportion $\hat{r} > 0$ of them that are actually bigger than M.

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

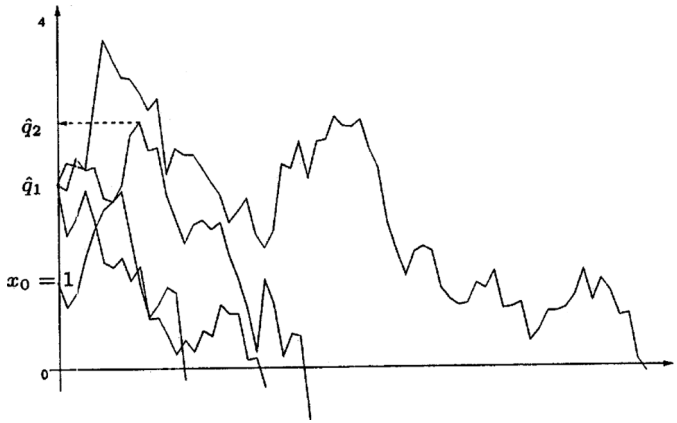


Figure 3. The second step of the algorithm, with $n = 4$ and $k = 1$.

- Compute the probability estimate. If we denote $p = \frac{k}{n}$, then the estimate of the rare event is simply

$$\hat{\alpha}_n = \hat{r} p^{\widehat{N}}.$$

3. CONSISTENCY

Thanks to the strong Markov property of the process and the continuity of its trajectories, at the end of step l, it is clear that, given $\hat{q}_{l-1}$, the random variables $(S^l_{n,j})_{1 \le j \le n}$ are i.i.d. according to the following law :

$$S^l_{n,j} \sim \mathscr{D}\Big(\sup_{0 \le t \le T_0} X_t^{x_0} \,\Big|\, \sup_{0 \le t \le T_0} X_t^{x_0} \ge \hat{q}_{l-1} \Big),$$

with the convention that $\hat{q}_0 = x_0$. We can write it a little bit simpler:

$$S^l_{n,j} \sim \mathscr{D}\Big(\sup_{0 \le t \le T_0} X_t^{\hat{q}_{l-1}} \,|\, \hat{q}_{l-1} \Big).$$

Let us also denote $S = \sup_{0 \le t \le T_0} X_t^{x_0}$, and F its distribution function. We suppose that F is continuous. (Note that this property is not implied by the continuity of trajectories.) We now define $\mathbf{F}$ a deterministic function of two variables (q_1, q_2) as follows:

$$\mathbf{F}(q_1, q_2) = \mathbb{P}(S \le q_2 \,|\, S > q_1).$$

Since F is continuous, we have the obvious identity

$$\mathbf{F}(q_1, q_2) = \frac{F(q_2) - F(q_1)}{1 - F(q_1)}.$$

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

Thus, each $S^l_{n,j}$ has the distribution function $\mathbf{F}(\hat{q}_{l-1}, .)$. Note that at each step, the algorithm chooses $\hat{q}_l$ so that $\mathbf{F}(\hat{q}_{l-1}, \hat{q}_l)$ is close to $q = (1-p)$. Since F is continuous, the random variables $U^l_{n,1}, \ldots, U^l_{n,n}$, with

$$U^l_{n,j} = \mathbf{F}\big(\hat{q}_{l-1}, S^l_{n,j}\big),$$

are identically distributed with uniform law on $[0, 1]$ and row-wise independent. In fact, the proof of consistency mainly relies on this simple argument. Before stating the theorem, we sum up the assumptions on our model.

Hypothesis ($\mathscr{H}$). The strongly Markov process $(X_t)_{t\geq 0}$ starts from $x_0 > 0$, with 0 as an attractive point. $(X_t)_{t\geq 0}$ has time-continuous trajectories and the distribution function F of the random variable $S = \sup_{0\leq t\leq T_0} X^{x_0}_t$ is continuous.

Theorem 1. *Under assumption ($\mathscr{H}$), we have*

$$\hat{\alpha}_n \xrightarrow[n\to\infty]{a.s.} \alpha$$

In order to prove this theorem, will need two simple lemmas.

Lemma 1. *Let $(U_{n,j})_{1\leq j\leq n}$ be a triangular array of identically distributed random variables with uniform law on $[0, 1]$ and row-wise independent. Let $k_n = \lfloor np \rfloor$, then*

$$U_{n,(n-k_n)} \xrightarrow[n\to\infty]{a.s.} q = 1 - p.$$

Proof. Let us denote by $\mathbb{G}_n$ the empirical distribution function of $(U_{n,j})_{1\leq j\leq n}$, and G the distribution function of the uniform law on $[0, 1]$, that is the identity on $[0, 1]$. First note the basic identity $\|\mathbb{G}_n - G\|_\infty = \|\mathbb{G}_n^{-1} - G\|_\infty$ (see [17], p. 86), and $U_{n,(n-k_n)} = \mathbb{G}_n^{-1}\big(1 - \frac{k_n}{n}\big)$. We have also

$$|U_{n,(n-k_n)} - q| \leq \big\|\mathbb{G}_n^{-1} - G\big\|_\infty + |p - k_n/n|.$$

Using Dvoretsky–Kiefer–Wolfowitz (DKW) inequality (see [15, 18]) we have

$$\mathbb{P}\left(\|\mathbb{G}_n - G\|_\infty > \frac{1}{\log n} - |q - k_n/n|\right) \leq 2\exp\left[-2n\left(\frac{1}{\log n} - |p - k_n/n|\right)^2\right],$$

which implies that

$$\mathbb{P}\left(|U_{n,(n-k_n)} - q| > \frac{1}{\log n}\right) \leq 2\exp\left[-2n\left(\frac{1}{\log n} - |p - k_n/n|\right)^2\right].$$

We conclude using Borel-Cantelli lemma and the convergence of the series on the right-hand-side, having noted that $|p - k_n/n| \leq \frac{1}{n}$.

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

Lemma 2. *Let $(V_{n,j})_{1\le j\le n}$ be a triangular array of random variables row-wise independent and identically distributed. Let H_n be the distribution function of the nth row, and $\mathbb{H}_n$ its empirical distribution function. Then*

$$\|H_n - \mathbb{H}_n\|_\infty \xrightarrow[n\to\infty]{a.s.} 0.$$

Proof. This is again a simple application of DKW inequality, since

$$\mathbb{P}\left(\|\mathbb{H}_n - H_n\|_\infty > \frac{1}{\log n}\right) \le 2\exp\left[-2n\left(\frac{1}{\log n}\right)^2\right].$$

We conclude using Borel–Cantelli lemma and the convergence of the series on the right-hand-side. □

Proof of the Theorem. First of all, we shall see that for all l :

$$\mathbf{F}(\hat{q}_{l-1}, \hat{q}_l) \xrightarrow[n\to\infty]{a.s.} 1-p. \tag{1}$$

This property is a direct application of Lemma 1.

Then, we should notice that

$$\prod_{k=1}^{l}(1-\mathbf{F}(\hat{q}_{k-1},\hat{q}_k)) = \prod_{k=1}^{l}\frac{1-F(\hat{q}_k)}{1-F(\hat{q}_{k-1})} = 1-F(\hat{q}_l) = \mathbb{P}\left(\sup_{0\le t\le T_0} X_t^{x_0} > \hat{q}_l \mid \hat{q}_l\right).$$

Next we will show that for all l

$$\left|\prod_{k=1}^{l}(1-\mathbf{F}(\hat{q}_{k-1},\hat{q}_k)) - p^l\right| \xrightarrow[n\to\infty]{a.s.} 0.$$

We already have the convergence for $l=1$. Then we proceed by induction. Assume that the previous convergence is true for some $l \ge 1$. Then we use that the product of two random variables will a.s. converge to the product of the limits.

Now we focus on the last step. Assume first that $\frac{\log\alpha}{\log p}$ is not an integer. Let $N = \lfloor\frac{\log\alpha}{\log p}\rfloor$. Then we have that a.s. for n large enough,

$$\prod_{k=1}^{N+1}[1-\mathbf{F}(\hat{q}_{k-1},\hat{q}_k)] < \alpha < \prod_{k=1}^{N}[1-\mathbf{F}(\hat{q}_{k-1},\hat{q}_k)],$$

that is,

$$1-F(\hat{q}_{N+1}) < \alpha < 1-F(\hat{q}_N),$$

which implies that

$$\hat{q}_N < M < \hat{q}_{N+1},$$

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

so that, a.s. for n large enough, the algorithm stops after $\widehat{N} = N$ iterations. Let us denote by $\mathbb{F}_{n,N}$ the empirical distribution function of the $\{S_{n,j}^N, 1 \leq j \leq n\}$. Using Lemma 2, we have that a.s.

$$|\mathbf{F}(\hat{q}_N, M) - \mathbb{F}_{n,N}(M)| \leq \|\mathbf{F}(\hat{q}_N, .) - \mathbb{F}_{n,N}\|_\infty \xrightarrow[n\to\infty]{a.s.} 0.$$

On the other hand, using the definition of N, we have that

$$\prod_{k=1}^{N} [1 - \mathbf{F}(\hat{q}_{k-1}, \hat{q}_k)] \cdot [1 - \mathbf{F}(\hat{q}_N, M)] = 1 - F(M) = \alpha,$$

which implies:

$$\lim_{n\to+\infty} [1 - \mathbb{F}_{n,N}(M)] = \frac{\alpha}{p^N} \quad \text{a.s.}$$

So we get

$$\lim_{n\to+\infty} \hat{\alpha}_n = \lim_{n\to+\infty} p^N [1 - \mathbb{F}_{n,N}(M)] = p^N \frac{\alpha}{p^N} = \alpha,$$

which gives the estimate consistency.

Finally, we consider the case when $\frac{\log \alpha}{\log p}$ is an integer. Again we set $N = \frac{\log \alpha}{\log p}$. In this case, using the same arguments, we have

$$\prod_{k=1}^{N} [1 - \mathbf{F}(\hat{q}_{k-1}, \hat{q}_k)] \xrightarrow[n\to\infty]{a.s.} p^N = \alpha.$$

So, for n large enough, the algorithm stops after $\widehat{N} = N$ or $\widehat{N} = (N+1)$ steps. We have to consider two cases: either $\hat{q}_N \geq M$, or $\hat{q}_N < M$, and the estimate may be written as

$$\hat{\alpha}_n = p^{N-1}[1 - \mathbb{F}_{n,N}(M)]\mathbb{1}_{\{\hat{q}_N \geq M\}} + p^N [1 - \mathbb{F}_{n,N+1}(M)]\mathbb{1}_{\{\hat{q}_N < M\}}.$$

We also have, using Lemma 2 in the same way as we did in the first part of the proof:

$$\lim_{n\to+\infty} [1 - \mathbb{F}_{n,N-1}(M)] = p,$$

and

$$\lim_{n\to+\infty} [1 - \mathbb{F}_{n,N}(M)] = 1.$$

Then we get

$$\begin{aligned} |\hat{\alpha}_n - \alpha| \leq & \left| p^{N-1}[1 - \mathbb{F}_{n,N}(M)] - p^N \right| \cdot \mathbb{1}_{\{\hat{q}_N \geq M\}} \\ & + \left| p^N [1 - \mathbb{F}_{n,N+1}(M)] - p^N \right| \cdot \mathbb{1}_{\{\hat{q}_N < M\}}, \end{aligned}$$

where both terms tend a.s. to 0, which concludes the proof. □

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

Remark. If we suppose that F^{-1} is continuous, the proof is shorter, since we do not need then Lemmas 1 and 2. Indeed, in that case, classical results say that empirical quantiles converge almost surely towards the true quantiles (see for example [18], Lemma 21.2, p. 305).

4. ASYMPTOTIC NORMALITY

Now we are interested in the variance of this estimation. Let us denote $r = \alpha p^{-N}$, so that r verifies $p < r \leq 1$.

Theorem 2. *Under the assumptions we have made so far, we have*

$$\sqrt{n}(\alpha - \hat{\alpha}_n) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

with

$$\sigma^2 = \alpha^2\left(N\frac{1-p}{p} + \frac{1-r}{r}\right).$$

In order to prove this theorem, we will need the following lemmas.

Lemma 3. *Let $(U_{n,j})_{1\leq j\leq n}$ be a triangular array of identically distributed random variables with uniform law on $[0,1]$ and row-wise independent. Let $k_n = \lfloor np \rfloor$, then*

$$\sqrt{n}(U_{n,(n-k_n)} - q) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, p(1-p)).$$

Proof. This is a direct application of [1] (Theorem 8.5.1, p. 223) to $\mathcal{U}(0,1)$-distributed i.i.d. random variables. □

We recall now a very classical probabilistic result.

Lemma 4. *Let $\{V_n, n \in \mathbb{N}\}$ and $\{W_n, n \in \mathbb{N}\}$ be two sequences of random variables such that*

$$V_n \xrightarrow[n\to\infty]{\mathcal{D}} V \quad \text{and} \quad W_n \xrightarrow[n\to\infty]{P} 0,$$

Then

$$V_n + W_n \xrightarrow[n\to\infty]{\mathcal{D}} V.$$

and

$$V_n W_n \xrightarrow[n\to\infty]{P} 0.$$

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

The following lemma, intuitively clear, is very useful in practice: roughly speaking, if we are only interested in the order statistics of the supremum, everything happens as if we had uniform and independent random variables.

Lemma 5. *Let $l \geq 1$, $\hat{q}_l$, $\hat{q}_{l+1}$, $\mathbf{F}$, and k_n be as before. For any test function $\varphi : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}[\varphi(\mathbf{F}(\hat{q}_l, \hat{q}_{l+1})) \mid \hat{q}_l] = \mathbb{E}[\varphi(U_{n,(n-k_n)})],$$

with $(U_{n,j})_{1\leq j\leq n}$ a triangular array of identically distributed random variables with uniform law on $[0, 1]$ and row-wise independent.

The last lemma is a little bit more technical.

Lemma 6. *Let (r_n) be a sequence of random variables and r a deterministic constant such that*

$$\forall n \in \mathbb{N},\ r_n \in (0, 1), \quad \text{and} \quad r \in (0, 1), \quad r_n \xrightarrow[n\to\infty]{a.s.} r,$$

and

$$\sqrt{n}(r_n - r) \xrightarrow[n\to\infty]{\mathscr{D}} \mathcal{N}(0, \sigma^2),$$

Let us consider next a triangular array $(B_{n,j})_{1\leq j\leq n}$ of random variables, with the nth row being conditionally to ρ_n i.i.d. Bernoulli trials, of parameter r_n (i.e., for all $1 \leq j \leq n$, $\mathbb{P}(B_{nj} = 1 \mid r_n) = r_n = 1 - \mathbb{P}(B_{nj} = 0 \mid r_n)$). Then we have the following result:

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} B_{nj} - r\right) \xrightarrow[n\to\infty]{\mathscr{D}} \mathcal{N}(0, s^2)$$

with $s^2 = \sigma^2 + r(1 - r)$.

Proof. In the sequel the notation $o(\cdot)$ refers to a.s. convergence. Let us first consider the conditional characteristic function

$$\begin{aligned}
\phi_{n,r_n}(t) &= \mathbb{E}\left[\exp\left(it\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} B_{nj} - r\right)\right) \,\middle|\, r_n\right] \\
&= e^{-itr\sqrt{n}}\left(r_n e^{\frac{it}{\sqrt{n}}} + (1 - r_n)\right)^n \\
&= \exp\left[-itr\sqrt{n} + n\log\left(r_n e^{\frac{it}{\sqrt{n}}} + (1 - r_n)\right)\right] \\
&= \exp\left[-itr\sqrt{n} + n\log\left(r_n\left(1 + \frac{it}{\sqrt{n}} - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right) + 1 - r_n\right)\right] \\
&= \exp\left[-itr\sqrt{n} + n\log\left(1 + \frac{itr_n}{\sqrt{n}} - \frac{r_n t^2}{2n} + o\left(\frac{1}{n}\right)\right)\right].
\end{aligned}$$

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

Then taking the expectation and developing the log,

$$\begin{aligned}
&\mathbb{E}[\phi_{n,r_n}(t)] \\
&\quad = \mathbb{E}\left[\exp\left[-itr\sqrt{n} + n\left(\left(\frac{itr_n}{\sqrt{n}} - \frac{r_n t^2}{2n}\right) - \frac{1}{2}\left(\frac{itr_n}{\sqrt{n}} - \frac{r_n t^2}{2n}\right)^2 + o\left(\frac{1}{n}\right)\right)\right]\right] \\
&\quad = \mathbb{E}\left[\exp\left[it\sqrt{n}(r_n - r) - \frac{1}{2}r_n(1-r_n)t^2 + o(1)\right]\right],
\end{aligned}$$

where

$$\frac{1}{2}r_n(1-r_n)t^2 + o(1) \xrightarrow[n\to\infty]{a.s.} \frac{1}{2}r(1-r)t^2,$$

and

$$\sqrt{n}(r_n - r) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

from which we get the convergence of the pair and then

$$\mathbb{E}[\phi_{n,r_n}(t)] \xrightarrow[n\to\infty]{} \exp\left[-\frac{1}{2}\sigma^2 t^2 - \frac{1}{2}r(1-r)t^2\right].$$

The limit is the characteristic function of a random variable of law $\mathcal{N}(0, \sigma^2 + r(1-r))$, which concludes the proof of the lemma.

Now we are able to prove the theorem.

Proof of the Theorem. We proceed like in the proof of Theorem 1. We begin by seeing that for all l

$$\sqrt{n}(1 - \mathbf{F}(\hat{q}_i, \hat{q}_{i+1}) - p) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}(0, p(1-p)).$$

This property is a direct application of Lemmas 3 and 5. We will use the following identity:

$$ab - cd = (a-c)(b-d) + (a-c)d + (b-d)c.$$

Now we proceed by induction:

$$\begin{aligned}
&\sqrt{n}\left(\prod_{k=1}^{l+1}\left[1 - \mathbf{F}(\hat{q}_{k-1}, \hat{q}_k)\right] - p^{l+1}\right) \\
&\quad = \sqrt{n}\left(\prod_{k=1}^{l}\left[1 - \mathbf{F}(\hat{q}_{k-1}, \hat{q}_k)\right] - p^{l}\right)(1 - \mathbf{F}(\hat{q}_l, \hat{q}_{l+1}) - p) \\
&\qquad + p\sqrt{n}\left(\prod_{k=1}^{l}\left[1 - \mathbf{F}(\hat{q}_{k-1}, \hat{q}_k)\right] - p^{l}\right) + p^l\sqrt{n}\left(1 - \mathbf{F}(\hat{q}_l, \hat{q}_{l+1}) - p\right). \quad (2)
\end{aligned}$$

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

For the first term, using Equation (1), we know that:

$$1 - \mathbf{F}(\hat{q}_l, \hat{q}_{l+1}) - p \xrightarrow[n\to+\infty]{a.s.} 0,$$

and by induction hypothesis

$$\sqrt{n}\left(\prod_{k=1}^{l} [1 - \mathbf{F}(\hat{q}_{k-1}, \hat{q}_k)] - p^l \right) \xrightarrow[n\to+\infty]{\mathscr{D}} \mathcal{N}(0, \sigma_l^2).$$

So that, thanks to Lemma 4, we have

$$\sqrt{n}\left(\prod_{k=1}^{l} \left[1 - \mathbf{F}(\hat{q}_{k-1}, \hat{q}_k)\right] - p^l \right)\left(1 - \mathbf{F}(\hat{q}_l, \hat{q}_{l+1}) - p\right) \xrightarrow[n\to+\infty]{P} 0.$$

We want to prove that the other terms in Equation (2) both converge in distribution. For this we use the characteristic function.

$$\phi_n(t) = \mathbb{E}\Bigg[\exp\Bigg(it\Bigg(p\sqrt{n}\Bigg(\prod_{k=1}^{l} \left[1 - \mathbf{F}(\hat{q}_{k-1}, \hat{q}_k)\right] - p^l \Bigg) + p^l\sqrt{n}\left(1 - \mathbf{F}(\hat{q}_l, \hat{q}_{l+1}) - p\right)\Bigg)\Bigg)\Bigg].$$

Thanks to the strong Markov property of the process:

$$\begin{aligned}\phi_n(t) &= \mathbb{E}\Bigg[\exp\Bigg(itp\sqrt{n}\Bigg(\prod_{k=1}^{l} \left[1 - \mathbf{F}(\hat{q}_{k-1}, \hat{q}_k)\right] - p^l \Bigg)\Bigg) \\ &\qquad \times \mathbb{E}\left[\exp\left(itp^l\sqrt{n}(1 - \mathbf{F}(\hat{q}_l, \hat{q}_{l+1}) - p)\right) \mid \hat{q}_1, \ldots, \hat{q}_l \right]\Bigg] \\ &= \mathbb{E}\Bigg[\exp\Bigg(itp\sqrt{n}\Bigg(\prod_{k=1}^{l} [1 - \mathbf{F}(\hat{q}_{k-1}, \hat{q}_k)] - p^l \Bigg)\Bigg) \\ &\qquad \times \mathbb{E}\left[\exp\left(itp^l\sqrt{n}(1 - \mathbf{F}(\hat{q}_l, \hat{q}_{l+1}) - p)\right) \mid \hat{q}_l \right]\Bigg].\end{aligned}$$

Lemma 5 ensures that we can write the last term in another way:

$$\mathbb{E}[\exp(itp^l\sqrt{n}\left(1 - \mathbf{F}(\hat{q}_l, \hat{q}_{l+1}) - p\right)) \mid \hat{q}_l] = \mathbb{E}\left[\exp\left(itp^l\sqrt{n}(1 - U_{n,(n-k_n)})\right)\right],$$

which is a deterministic complex number. Thus,

$$\begin{aligned}\phi_n(t) = \mathbb{E}\Bigg[\exp\Bigg(itp\sqrt{n}\Bigg(\prod_{k=1}^{l} \left[1 - \mathbf{F}(\hat{q}_{k-1}, \hat{q}_k)\right] - p^l \Bigg)\Bigg)\Bigg] \\ \times \mathbb{E}\left[\exp\left(itp^l\sqrt{n}(1 - U_{n,(n-k_n)})\right)\right].\end{aligned}$$

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

And now we just have to show that both terms have a Gaussian limit. By induction hypothesis, we know that

$$p\sqrt{n}\left(\prod_{k=1}^{l}\left[1-\mathbf{F}(\hat{q}_{k-1},\hat{q}_k)\right]-p^l\right)\xrightarrow[n\to+\infty]{\mathscr{D}}\mathcal{N}\left(0,p^2\sigma_l^2\right).$$

By Lemma 3, we know that

$$p^l\sqrt{n}(1-U_{n,(n-k_n)})\xrightarrow[n\to+\infty]{\mathscr{D}}\mathcal{N}(0,p^{2l+1}(1-p)).$$

Thus,

$$p\sqrt{n}\left(\prod_{k=1}^{l}\left[1-\mathbf{F}\left(\hat{q}_{k-1},\hat{q}_k\right)\right]-p^l\right)+p^l\sqrt{n}\left(1-\mathbf{F}(\hat{q}_l,\hat{q}_{l+1})-p\right)$$
$$\xrightarrow[n\to+\infty]{\mathscr{D}}\mathcal{N}\left(0,p^2\sigma_l^2+p^{2l+1}(1-p)\right).$$

By Lemma 4, we conclude that

$$\sqrt{n}\left(\prod_{k=1}^{l+1}[1-\mathbf{F}(\hat{q}_{k-1},\hat{q}_k)]-p^{l+1}\right)\xrightarrow[n\to+\infty]{\mathscr{D}}\mathcal{N}(0,\sigma_{l+1}^2),$$

with $\sigma_{l+1}^2=p^2\sigma_l^2+p^{2l+1}(1-p)$. From this recursion, we deduce:

$$\sigma_N^2=Np^{2N-1}(1-p).$$

Now we deal with the last step. Let $\widehat{N}$ be the (random) number of steps of the algorithm. Let us first suppose that $\frac{\log\alpha}{\log p}$ is not an integer (we recall that $N=\lfloor\frac{\log\alpha}{\log p}\rfloor$) and assume that the algorithm is deterministically stopped after $(N+1)$ steps. Then the estimate is $\hat{\alpha}_{d,N}=p^N\hat{\alpha}_N$, where

$$\hat{\alpha}_N=\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}_{\{S_{n,j}^N\geq M\}}.$$

The variables $\mathbb{1}_{\{S_{n,j}^N\geq M\}}$ are i.i.d. Bernoulli trials, conditionally to $\hat{q}_N$. The parameter of the Bernoulli is

$$r_n=\mathbb{P}\left(\mathbb{1}_{\{S_{n,j}^N\geq M\}}=1\,|\,\hat{q}_N\right)=1-\mathbf{F}(\hat{q}_N,M)=\frac{\alpha}{1-F(\hat{q}_N)}. \tag{3}$$

We have already shown that

$$\sqrt{n}\left[\prod_{k=1}^{N}\left(1-\mathbf{F}(\hat{q}_{k-1},\hat{q}_k)\right)-p^N\right]\xrightarrow[n\to+\infty]{\mathscr{D}}\mathcal{N}(0,\sigma_N^2), \tag{4}$$

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

Then we write, using the definition of F and $\mathbf{F}$:

$$\prod_{k=1}^{N} \left(1 - \mathbf{F}(\hat{q}_{k-1}, \hat{q}_k)\right) = 1 - F(\hat{q}_N).$$

Let $r = \alpha p^{-N}$. Using Equation 3, we get:

$$\sqrt{n}(r_n - r) = \alpha\sqrt{n}\left[\frac{p^N - (1 - F(\hat{q}_N))}{p^N(1 - F(\hat{q}_N))}\right].$$

We know from the proof of Theorem 1 that

$$\frac{\alpha}{p^N(1 - F(\hat{q}_N))} \xrightarrow[n\to+\infty]{a.s.} \frac{\alpha}{p^N}\frac{1}{p^N}. \tag{5}$$

So we have that

$$\sqrt{n}(r_n - r) \xrightarrow[n\to+\infty]{\mathscr{D}} \mathcal{N}(0, \sigma_r^2),$$

with

$$\sigma_r^2 = \frac{\alpha^2}{p^{2N+1}} N(1 - p).$$

Then we apply Lemma 6 to get that

$$\sqrt{n}(\hat{\alpha}_{d,N} - \alpha) \xrightarrow[n\to+\infty]{\mathscr{D}} \mathcal{N}(0, \sigma^2),$$

with

$$\sigma^2 = \alpha^2\left(N\frac{1 - p}{p} + \frac{1 - r}{r}\right).$$

Then we come back to the true (random) $\widehat{N}$. Let us consider

$$\tilde{\alpha}_N = \hat{\alpha}\,\mathbb{1}_{\{\widehat{N}=N\}} + \gamma\,\mathbb{1}_{\{\widehat{N}\neq N\}},$$

where γ is a random variable whose law is the law of $\hat{\alpha}_{d,N}$, conditionally to $\widehat{N}$ and $\hat{q}_{\widehat{N}\wedge N}$. Then it is quite obvious to see that $\tilde{\alpha}_N$ has the same law as $\hat{\alpha}_{d,N}$, implying they have both the same asymptotic behavior in distribution. On the other hand we get for all $\varepsilon > 0$,

$$\mathbb{P}(|\sqrt{n}(\hat{\alpha} - \tilde{\alpha}_N)| > \varepsilon) \leq \mathbb{P}(|\hat{\alpha} - \tilde{\alpha}_N| > 0) = \mathbb{P}(\widehat{N} \neq N) \xrightarrow[n\to+\infty]{} 0,$$

as we saw in the proof of Theorem 1 that $\widehat{N} \xrightarrow[n\to+\infty]{a.s.} N$. Now from all this we conclude that $\sqrt{n}(\hat{\alpha} - \alpha)$ and $\sqrt{n}(\tilde{\alpha}_N - \alpha)$ both converge to the same limit in distribution.

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

If $N = \frac{\log \alpha}{\log p}$ is an integer, we combine the reasoning at the end of the proof of Theorem 1 (i.e., distinction of two cases) and, in each case, the calculus of variance above to obtain the asymptotic normality with variance:

$$\sigma^2 = \alpha^2 N \frac{1-p}{p}.$$

□

5. NUMERICAL EXAMPLE

We have implemented this algorithm with trajectories following a Brownian process with drift. Noting B_t a Brownian motion, the process studied is $X_t = B_t + \mu t$, with $\mu < 0$ (see Figure 4). The drift was taken to have a motion going quickly to 0. This process clearly satisfies assumption $\mathcal{H}$. Moreover, it is simple enough so that analytical results are well-known about it.

We note $H_{a,b} = H = \min\{s > 0 : X_s \notin [a, b]\}$. The expression of the probability of reaching b before a starting from $x \in [a, b]$ is given by [3]:

$$\mathbb{P}_x(X_H = b) = e^{\mu(b-x)} \frac{\sinh((x-a)|\mu|)}{\sinh((b-a)|\mu|)}$$

Let us compare this to our numerical results.

The first problem we had to solve was that we were considering a continuous process, which is impossible in computing. We had to choose a step δt and consider the process at every $k \times \delta t$. This step has to be

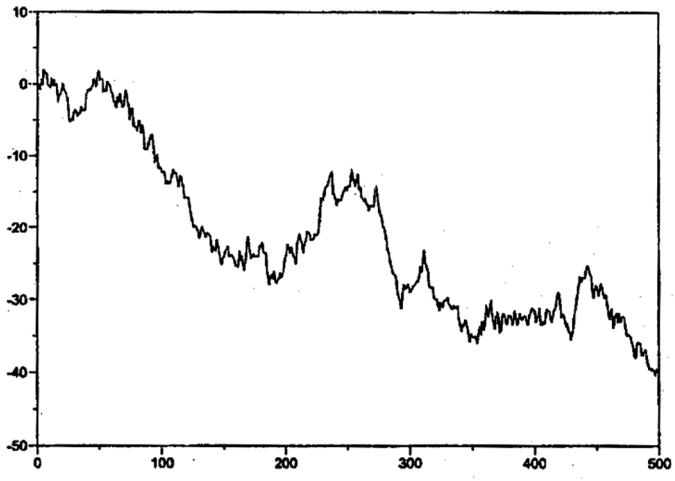


Figure 4. Brownian motion with negative drift starting from 0.

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

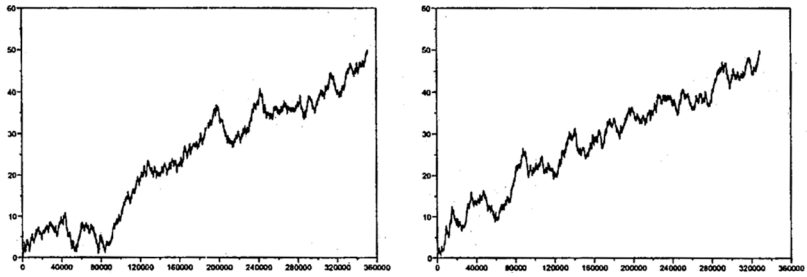


Figure 5. Two independent trajectories reaching 50 before 0 starting from 1.

small enough to avoid clipping the process, which could introduce a bias in the estimation. Examples of trajectories reaching the rare set are given in Figure 5.

We will illustrate first the a.s. convergence. We ran our algorithm on the above example with parameters $a = 0$, $b = 12$, $\mu = -1$, $x_0 = 1$, such that the rare event probability is $\alpha \approx 2.412 \times 10^{-10}$. Figure 6 gives the relative error as a function of the number n of particles. For $n = 20{,}000$ we have an error as low as 5%.

Then we illustrate the asymptotic normality. As we need to run the algorithm many times to estimate the law of the estimator, we chose a setting where α is not very small, but about 0.1244. Estimating of the same probability α 1,000 times gives the histogram of Figure 7. This confirms the fact that the distribution of the estimating values

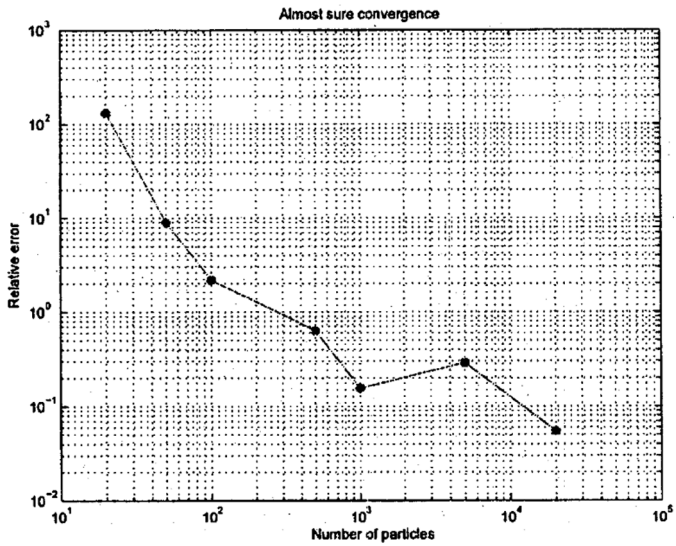


Figure 6. Almost sure convergence of the estimator as a function of n.

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

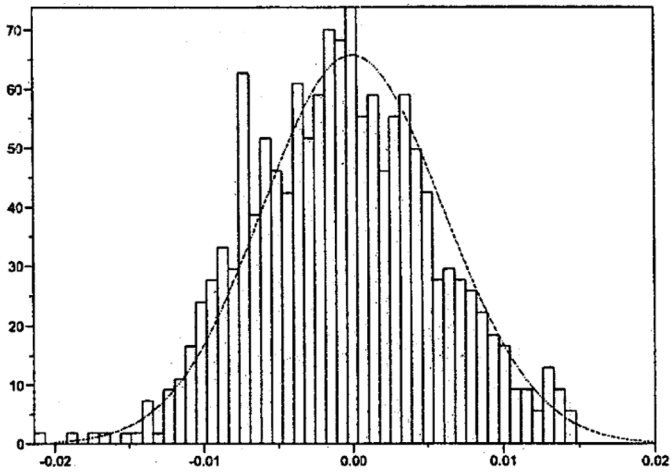


Figure 7. Histogram of the differences between γ and its estimations, with $n = 1,000$, for 1,000 instances. The curve represents the limit Gaussian distribution.

tends asymptotically towards a Gaussian distribution. Then we show the convergence of the variance multiplied by n on Figure 8. In this setting, we computed the asymptotic value which is about 0.0373, as shown by the dashed line on the figure. Here we made 250 runs for each value of n.

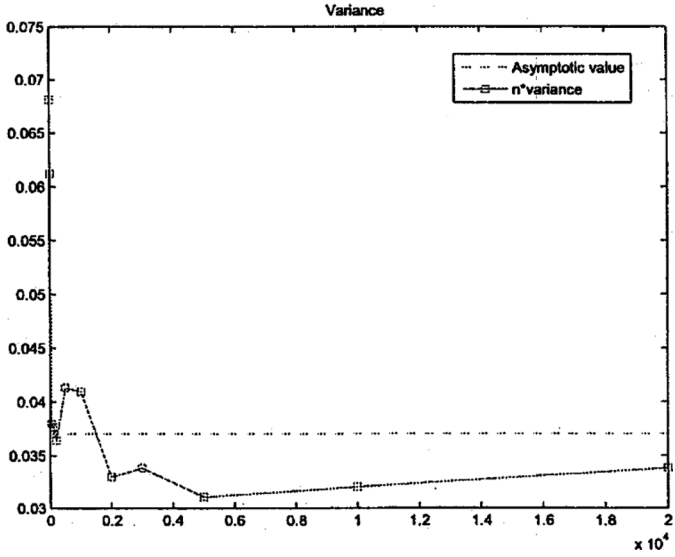


Figure 8. $n \times$ variance of the estimator as a function of n.

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

Once this algorithm is implemented, it can be improved to give more information than only estimating probabilities. The first idea was to keep the trajectories hitting the rare event in order to know what succession of events has reached the event we are interested in.

The thing is to keep not only the maxima of the trajectories but all the trajectories reaching the thresholds. In one dimension, there are no problems with knowing where the trajectory comes from—they do not have any choice. The number of different trajectories that can be created with this algorithm is $k^{\widehat{N}-1}\hat{r}n$, even if we can create only k totally independent.

The probability of a Brownian motion with drift -0.02 starting from 1 and reaching 50 before 0 is 1.014×10^{-9}. In a reasonable time (half a dozen minutes), we can disclose a few trajectories reaching 50 starting from 1.

Then, after keeping the whole trajectories, we can study the exit time compared to the theory given by [3] (with the same notation):

$$\mathbb{P}_x(H \in dt, X_H = b) = e^{\mu(b-x)-\frac{\mu^2 t}{2}} ss_{x-a,b-a}(t)\, dt$$

with

$$ss_{u,v}(t) = \sum_{k=-\infty}^{\infty} \frac{v-u+2kv}{\sqrt{2\pi}\, t^{3/2}} e^{-\frac{(v-u+2kv)^2}{2t}}.$$

The convergence of the infinite sum is extremely fast, thereby we can truncate it and calculate the density of the exit time in b. Moreover we are able of calculating many trajectories reaching b—which are not altogether independent. So we have numerical estimations of the exit time in b. We draw the histogram of these values and compare it to the theory. We can see on Figure 9 that the numerical results agree with the theory, which means that the adaptive algorithm gives a good sample of trajectories reaching b.

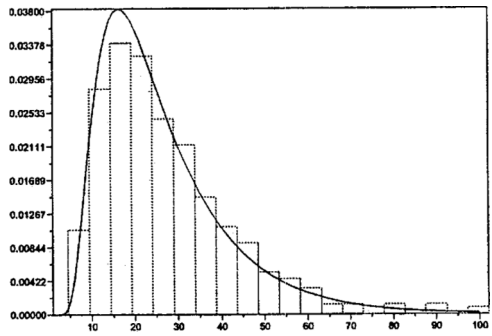


Figure 9. Theoretical curve and histogram of the numerical values for the exit time in b.

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

6. COMPARISON WITH EXISTING ALGORITHMS

6.1. Complexity and Variance

Now we want to have an idea about the complexity of the algorithm. To this aim let us consider the following simplification: We suppose that the simulation of a trajectory $\left(X_t^x\right)_{0\le t\le T_0}$, between its starting point $x > 0$ and the first time T_0 it hits 0, has a cost C which is approximately constant, independent of the initial condition x.

For the first step, with the assumption above, the cost for simulating n trajectories is linear in n. The finding of the maxima of the n trajectories is also linear in n. The sorting of these maxima is, in expectation, in $O(n \log n)$. Thus the first step of the algorithm has an expected complexity in $O(n \log n)$. For n large enough, there is a finite number of steps, that is $(N+1)$, with $N = \lfloor \frac{\log \alpha}{\log p} \rfloor$. Finally the total cost of the algorithm is in $O(n \log n)$ operations.

We are now interested in the precision of the estimator. We have seen above that the variance of $\hat{\alpha}$ is

$$\sigma^2 = \alpha^2 \left(N \frac{1-p}{p} + \frac{1-r}{r} \right),$$

with $p < r < 1$. If for simplicity we suppose that $\alpha = p^N$, then the normalized variance is

$$\frac{\sigma^2}{\alpha^2} = N \frac{1-p}{p} \approx \frac{\log \alpha}{\log p} \cdot \frac{1-p}{p}.$$

In real life applications, the only parameter that is fixed a priori is the small probability α to estimate. So the question is: What is the optimal choice for p? A straightforward study of the variance for p varying between α and 1 proves that σ^2 decreases when p goes to one.

This result is intuitively clear and merely says that if we want a precise estimate for α, we just have to put a lot of intermediate levels. But, of course, the complexity of the algorithm is then increasing since the number of levels is $\frac{\log \alpha}{\log p}$. So the choice of the parameter p, or equivalently the choice of the number N of levels, depends on what we want: a precise estimate or a quick algorithm.

6.2. Classical Multilevel Splitting

A usual way for estimating rare event probability is the multilevel splitting algorithm. The splitting idea is widespread in Monte Carlo methods (see for instance [11], p. 131). Its application to rare event estimation is first due to Kahn and Harris in the setting of

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

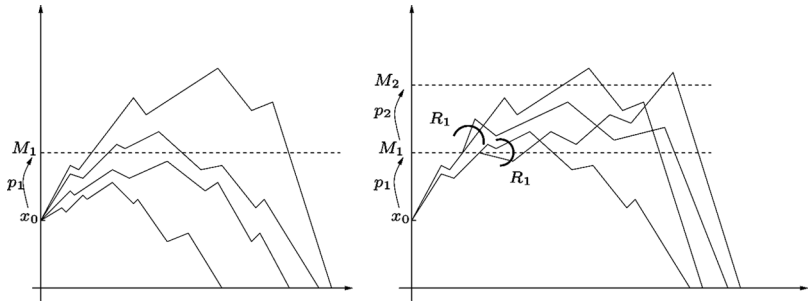


Figure 10. Multilevel splitting as seen by Lagnoux, with $n = 4$ and $R_1 = 2$.

particle transmission [12]. In 1970, Bayes proposed to apply it in the field of waiting queues [2].[1] This idea was rediscovered 20 years later by Villén-Altamirano and Villén-Altamirano [19]: This is what they call the RESTART method (REpetitive Simulation Trials After Reaching Thresholds) with applications in telecommunication traffic and reliability. In a simplified context, the idea has been theoretically studied by Glasserman et al. in several papers [8–10], and more recently by Lagnoux [13]. We refer the reader to [10] for a precise discussion and many references about splitting algorithms and their applications.

We can describe the algorithm in the simplified form of these last authors on our Markov process example. Let us denote A the event "reaching M before 0, starting from x_0," then consider $A = A_N \subset A_{N-1} \subset \cdots \subset A_1$ an increasing sequence of events. For us, this is equivalent to considering a decreasing sequence of levels $M = M_N > M_{N-1} > \cdots > M_1 > x_0$, each M_i being the threshold between A_{i-1} and A_i. Let $p_1 = \mathbb{P}(A_1)$ and $p_{i+1} = \mathbb{P}(A_{i+1} \mid A_i)$. These probabilities are bigger than $\alpha = \mathbb{P}(A)$ and thus easier to estimate. Moreover, since the process is Markov, the following product decomposition holds:

$$\alpha = p_1 p_2 \dots p_N.$$

In the version of Lagnoux, the idea of splitting is to simulate n paths starting from x_0, to duplicate R_1 times those who have reached M_1 before 0 (which happens with probability p_1), then to duplicate R_2 times those who have reached M_2 before 0 starting from M_1 (which happens with probability p_2), etc. (see Figure 10).

An unbiased estimator of α is

$$\tilde{\alpha} = \frac{n_A}{nR_1 \dots R_N},$$

[1]Unfortunately, he introduces some confusion in his paper by using the term "importance sampling" for what generally is named "importance splitting."

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

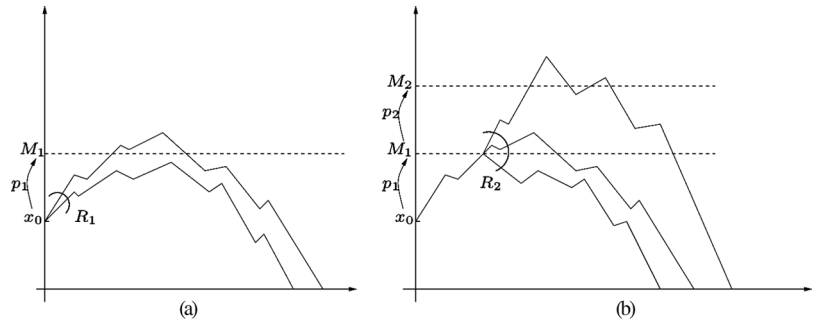


Figure 11. Multilevel splitting as seen by Glasserman, with $R_1 = 2$ and $R_2 = 3$.

where n_A is the number of trajectories that have reached M before 0. The complexity of the algorithm is this time in $O(n)$, which is less than for the adaptive multilevel splitting algorithm.

As Glasserman et al. noticed, one can also see the multilevel splitting method in terms of branching processes. Suppose there is initially one ancestor, his offspring has a binomial distribution $\mathcal{B}(R_1, p_1)$, each child has himself an offspring with binomial distribution $\mathcal{B}(R_2, p_2)$, etc. (see Figure 11). Compare to the version of Lagnoux, we see that it is the same idea, there is only a shift of one step in the splitting.

Suppose that if the number N and the levels $(M_1, \ldots, M_N)$ are fixed, we can compute the probabilities $(p_1, \ldots, p_N)$, and vice versa. Then, if the sequence $(R_1, \ldots, R_N)$ is fixed also, we can compute the variance of the estimator $\tilde{\alpha}$: either through direct calculus [13], or via the theory of branching processes [10]. The natural question that arises then is: What is the best choice for these sequences, in terms of estimator variance and complexity? Glasserman et al. [10] have shown that asymptotically when $\alpha \to 0$, the best thing we have to do is to take all the p_i's equal to $p = \alpha^{1/N}$ and all the R_i's equal to $R = 1/p$ (if this quantity is not an integer, just randomize the algorithm so that $\mathbb{E}[R] = 1/p$). Lagnoux has shown the same result without any asymptotics on α. In terms of branching processes, this result says that the best compromise between complexity and variance is reached in the critical regime. No surprise in this story: if $R > 1/p$, the variance is smaller than simple Monte Carlo, but the number of paths will explode,[2] and if $R < 1/p$, there is no substantial gain in estimation compared to naive Monte Carlo.

In relation with this last point, we could have described our algorithm in a slightly more general way, that means: with parameters

[2]In this case, most of the computing time is spent with highly correlated trajectories.

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

$(k_1, n_1), \ldots, (k_N, n_N)$ instead of the same couple (k, n) at each step. Anyway, when n and k go to infinity so that the ratio k/n goes to p, it is clear that we get closer and closer to the classical splitting algorithm with regular levels. Thus, the result obtained by preceding authors on the classical splitting algorithm shows that the best thing we have to do in our case is to keep the same couple (k, n) at each step.

For the sake of simplicity, let us suppose again that $\alpha = p^N$, with $R = \frac{1}{p}$ an integer. Then the variance s^2 of the estimator $\tilde{\alpha}$ in classical multilevel splitting is just the same[3] as for our $\hat{\alpha}$:

$$\frac{s^2}{\alpha^2} = N\frac{1-p}{p}.$$

The difference between the variances of $\hat{\alpha}$ and $\tilde{\alpha}$ is the following: Since the adaptive version needs interactions between the n trajectories, its variance is asymptotic when n goes to infinity. This is not the case with classical splitting, where the trajectories are independent. On the other hand, the above formula is only a best case for classical multilevel splitting (regular levels), whereas it is always asymptotically granted in the adaptive version, as long as k/n is kept fixed. In other words, in terms of estimators fluctuations, classical multilevel splitting will never perform better than adaptive multilevel splitting.

Indeed, in real life applications, it is unfortunately impossible to have regular levels: the systems are usually so complex that any analytic calculus about them is just out of question. In this context, our adaptive algorithm is very useful: the levels $(\hat{q}_1, \ldots, \hat{q}_N)$ are determined during the algorithm and they are in fact approximations of the true quantiles $(M_1, \ldots, M_N)$. So, at the expense of a multiplying factor $\log n$ in the complexity, we have an algorithm that is really suitable for applications.

Concerning the issue of fixing the levels, the authors discovered a posteriori the thesis of Garvels [6]. In Section 3.3.1., he proposes to estimate them during a trial run, in the same way as we do, and then to apply classical splitting. He writes: "Care has to be taken that enough samples are used determining the thresholds, otherwise the thresholds may become biased and force the real simulation in the wrong area. A good rule of thumb is to devote 10% of the simulation effort to a trial run which will determine all the thresholds." To our knowledge, he does not mention that both tasks can be treated simultaneously, with a consistency result and an "optimal" asymptotic variance.

[3]Note that the formula in [10] (p. 589) seems to be different, but this is only due to the first step of their algorithm.

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

6.3. Multilevel Splitting as a Feynmac–Kac Model

Another variant of multilevel splitting has been recently proposed by Del Moral [4, 5]. Like in classical multilevel splitting, it requires to fix the thresholds before beginning simulations (in other words, it is not adaptive). Anyway, one of the benefits of this version is to connect multilevel splitting techniques with interacting particle systems approximations of Feynman–Kac distributions. This last topic has been studied intensively since the beginning of the 1990s (see [4]), so that precise and general results can be transposed in the field of rare event analysis.

We can describe briefly the algorithm in the context of the above mentioned Markov process: suppose the number N and the levels $M_1 < \cdots < M_N = M$ are fixed, like in classical splitting. At time 0, the algorithm starts with n independent copies of the process X that are stopped when they reach 0 or M_1, whatever occurs the first. The particles which reach 0 before M_1 are killed and randomly redistributed among those having reached the first level, producing offsprings. If the whole system is killed, the algorithm is stopped. Otherwise, with the offprings, there are still n particles at the first level M_1. In a second step, the n particles in the first level evolve according to the same rule of the process X. Here, again, particles that reach 0 before M_2 are killed and for each killed one we randomly choose one of the particles that have reached the second level and add an offspring to it (see Figure 12). Then the process goes on to the next level, and so on until the rare event is reached.

For comparison, the cost is the same as for classical splitting, in $O(n)$. The asymptotic normality of the estimator is proved, with the same variance as before, i.e., $N \cdot \frac{1-p}{p}$ (by adapting for instance Theorem 12.2.2 in [4] in our framework).

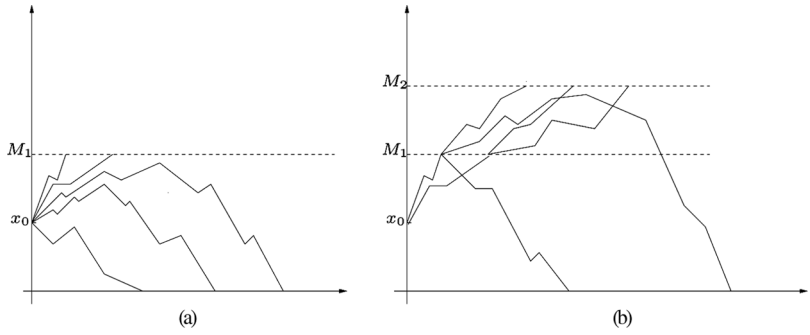


Figure 12. Multilevel splitting as a Feynmac–Kac model, with $n = 5$.

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

6.4. Generalization

The assumptions made to prove the consistency and the asymptotic normality of $\hat{\alpha}$ are, mainly: The process is Markov, the distribution function F of its supremum is continuous, the process is in one dimension.

Even if, for now, we do not have theoretical results in more general cases, the algorithm itself is quite versatile. In the following, we briefly present two examples where the above assumptions are not satisfied, but where first experimentations give promising results.

Examples.

- **Saturation of a waiting queue buffer.** Consider a queue with 0 as a positive recurrent state. In other words, the tasks are treated faster than the clients arrive, so that the system is stable. Suppose moreover that the system has a huge but finite capacity, so that the saturation will be a rare event. We want to estimate its probability with adaptive splitting.
  The process $(X_t)_{t\geq 0}$ is the number of clients in the system. Note first that in general this process is not Markov. Secondly, $(X_t)_{t\geq 0}$ is not continuous. Consequently, we cannot apply directly our theoretical results. Nevertheless, there is no problem in adapting our algorithm in this context. The first steps are described in Figure 13.
- **Self avoiding walks.** Let us now mention a two-dimensional problem. We consider self-avoiding walks (SAW) in $\mathbb{Z}^2$, which serves as a model for molecular chains. We want to calculate the probability of having long chains and draw some of them. A SAW is simply a random walk on $\mathbb{Z}^2$, conditioned on the fact that it never visited twice the same site. The proportion of such trajectories upon all possible chains decreases very fast with its length n, making it a very rare event. See [14] for a monograph on SAW.

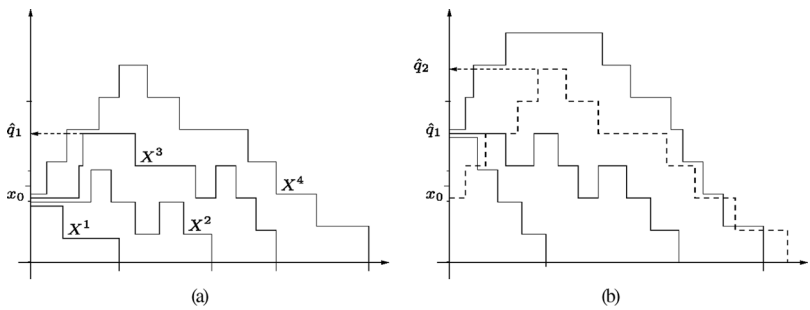


Figure 13. Adaptive algorithm for a waiting queue, with $n = 3$ and $k = 1$.

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

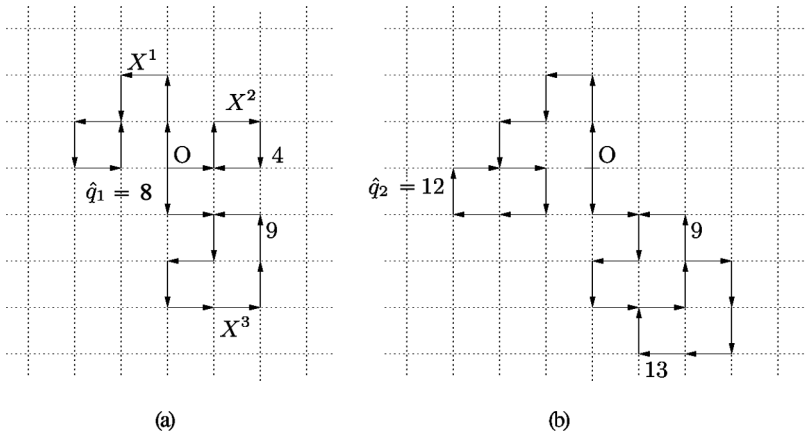


Figure 14. Adaptive algorithm for self-avoiding walks, with $n = 3$ and $k = 1$.

Here again, even if theory is only established for one-dimensional problems, adapting the algorithm is straightforward. The algorithm is described in Figure 14. The supremum of the trajectory in previous examples is just replaced by the length of the random walk until it hits himself. Note that in general, for multidimensional problems, this is the difficult question: Which criterion does really measure the fact that we are approaching the rare event? That is what Garvels et al. call "the choice of the importance function" [7].

As in the historical paper [16], we focused on the extension of the chain. The extension of a SAW X of length ℓ is defined as

$$s(\ell) = \mathbb{E}[X(\ell)^2].$$

It is proved that the limit

$$\lim_{\ell \to +\infty} \frac{s(\ell)}{\ell^\gamma}$$

exists and is finite for some $\gamma > 0$. But in $\mathbb{Z}^2$ the exact value of γ is still unknown, and it is conjectured [14] that $\gamma = 3/2$. In Figure 15, we present estimates of $s(\ell)$ as a function of ℓ. These were computed for $\ell \leq 150$ with $n = 150{,}000$, and $k = 50{,}000$. To estimate γ we decided to remove the beginning because we are not yet in the asymptotic regime. Keeping the values for $50 \leq \ell \leq 150$, and fitting the logarithms with a simple linear regression, we found $\gamma \simeq 1.511$, which is compatible with the conjecture.

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

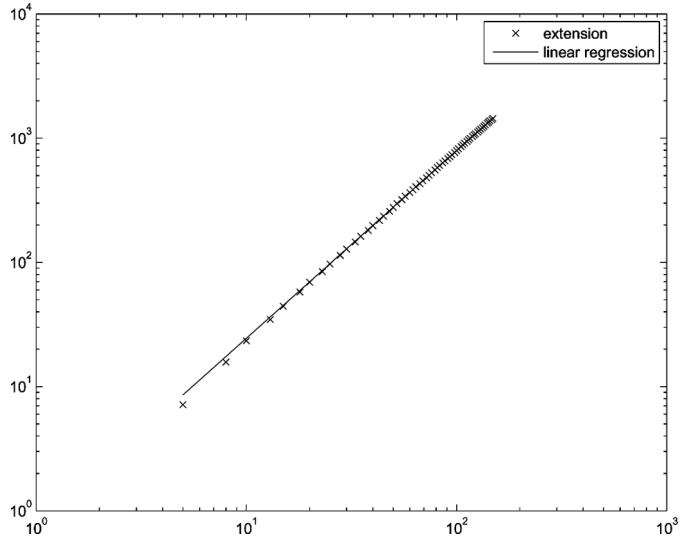


Figure 15. SAW extension.

7. CONCLUSION

We have presented a new algorithm for rare event analysis. This algorithm belongs to the multilevel family, but does not require the splitting levels to be set in advance. Instead it adapts them on the fly during the simulation. In the one dimensional case, we showed that there is no loss in variance compared to the other algorithms with optimal splitting levels, and only a slight increase in complexity.

REFERENCES

1. Arnold, B.C., Balakrishnan, N., and Nagaraja, H.N. 1992. *A First Course in Order Statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.
2. Bayes, A.J. 1970. Statistical techniques for simulation models. *Australian Comput. J.* 2:180–184.
3. Borodin, A.N., and Salminen, P. 1996. *Handbook of Brownian Motion – Facts End Formulae*. Birkhäuser Verlag, Basel, Boston, Berlin.
4. Del Moral, P. 2004. *Feynman–Kac Formulae, Genealogical and Interacting Particle Systems with Applications*. Probability and its Applications, Springer-Verlag, New York.
5. Del Moral, P., and Garnier, J. 2004. Genealogical particle analysis of rare events. Submitted to the *Annals of Applied Probability*. Available at http://www.lsp.ups-tlse.fr/Garnier/.

Downloaded By: [Cérou, Frédéric] At: 10:05 6 April 2007

6. Garvels, M.J.J. 2000. The Splitting Method in Rare Event Simulation. Thesis, University of Twente, Twente, May. Available at http://www.math.utwente.nl/dos/sor/AiO's/Garvels/Thesis.
7. Garvels, M.J.J., Kroese, D.P., and Van Ommeren, J.-K.C.W. 2002. On the importance function in splitting simulation. *European Transactions on Telecommunications* 13(4):363–371. Available at http://www.maths.uq.edu.au/kroese/.
8. Glasserman, P., Heidelberger, P., Shahabuddin, P., and Zajic, T. 1998. A large deviations perspective on the efficiency of multilevel splitting. *IEEE Trans. Automat. Control* 43(12):1666–1679.
9. Glasserman, P., Heidelberger, P., Shahabuddin, P., and Zajic, T. 2000. A look at multilevel splitting. In: *Monte Carlo and Quasi-Monte Carlo Methods 1996 (Salzburg)*. Volume 127 of Lecture Notes in Statist, Springer, New York, 98–108.
10. Glasserman, P., Heidelberger, P., Shahabuddin, P., and Zajic, T. 1999. Multilevel splitting for estimating rare event probabilities. *Oper. Res.* 47(4):585–600.
11. Hammersley, J.M., and Handscomb, D.C. 1965. *Monte Carlo Methods*. Methuen & Co. Ltd., London.
12. Kahn, H., and Harris, T.E. 1951. Estimation of particle transmission by random sampling. *National Bureau of Standards Appl. Math. Series* 12:27–30.
13. Lagnoux, A. 2005. Rare event simulation. *Probability in the Engineering and Informational Sciences* 20(1):45–66.
14. Neal Madras and Gordon Slade. 1993. *The Self-Avoiding Walk*. Probability and its Applications, Birkhäuser Boston Inc., Boston, MA.
15. Massart, P. 1990. The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Ann. Probab.* 18(3):1269–1283.
16. Rosenbluth, M.N., and Rosenbluth, A.W. Monte Carlo calculation of the average extension of molecular chains. *Journal of Chemical Physics* 23(2), fev.
17. Shorack, G.R., and Wellner, J.A. 1986. *Empirical Processes with Applications to Statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.
18. van der Vaart, A.W. 1998. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge.
19. Villén-Altamirano, M., and Villén-Altamirano, J. 1994. RESTART: a straightforward method for fast simulation of rare events. In: Proceedings of the 1994 Winter Simulation Conference. Jeffrey, D. Tew, Mani, S., Manivannan, Deborah A. Sadowski, and Andrew F. Seila, (Eds.), Orlando 1994, December 1994, 282–289.