Rare Event Simulation for the Stationary Distribution of a Markov Chain

Krzysztof Bisewski[1]

Joint work with:

Daan Crommelin[2,3], **Michel Mandjes**[3]

[1]University of Lausanne, [2]Centrum Wiskunde & Informatica,

[3]University of Amsterdam

PART I: Multilevel Splitting

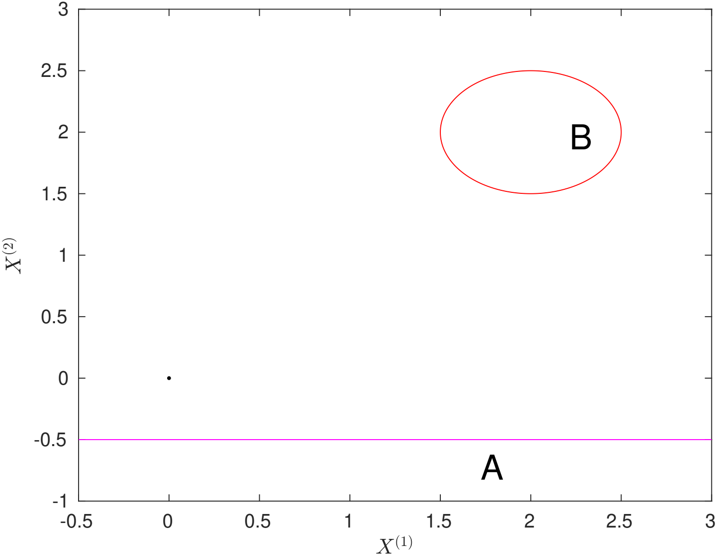
Multilevel Splitting setting

Let $(X_t)_{t\in[0,\infty)}$ be an $\mathbb{R}^d$-valued Markov process with initial condition $X_0 = \mathbf{0}$. We are interested in finding:

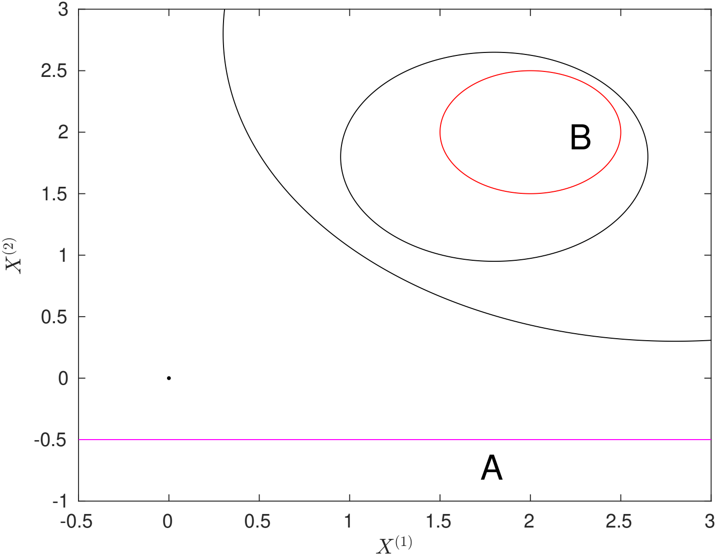
$$p := \mathbb{P}(\tau_B < \tau_A),$$

where $\tau_A := \inf\{t > 0 : X_t \in A\}$ and $\tau_B := \inf\{t > 0 : X_t \in B\}$. Here, B is very small and A is some large absorbing set.

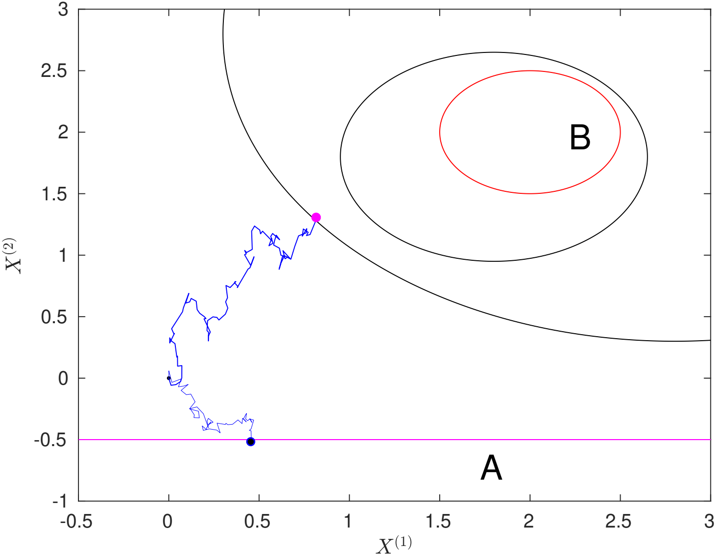
Multilevel Splitting



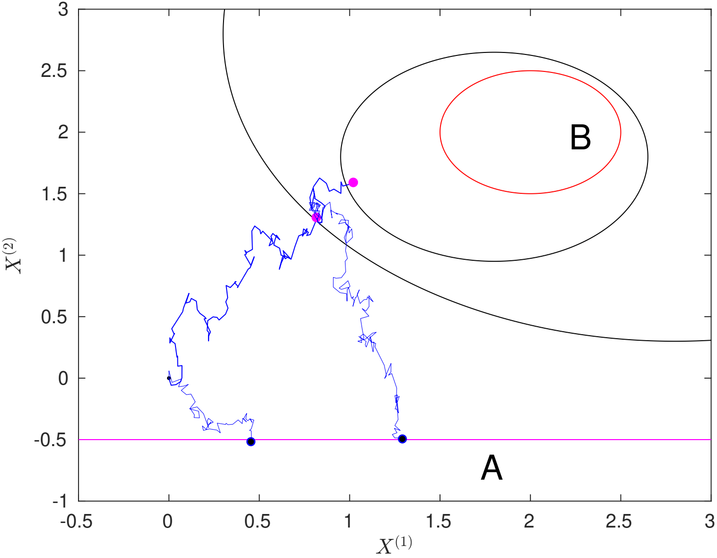
Multilevel Splitting



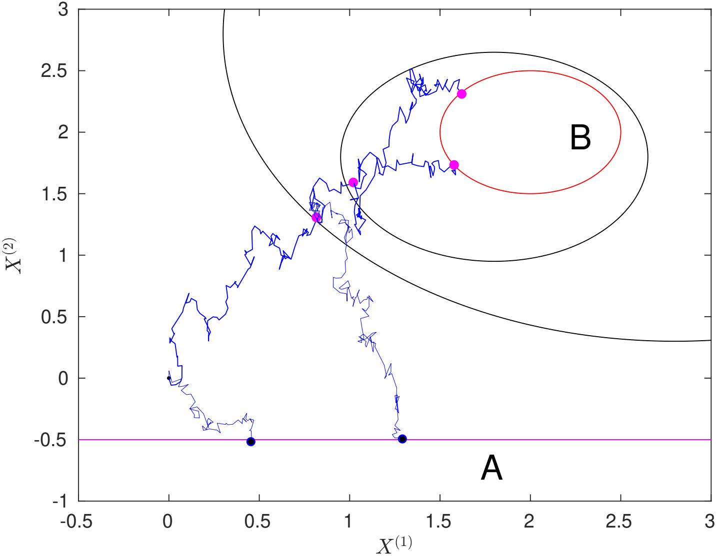
Multilevel Splitting



Multilevel Splitting



Multilevel Splitting



Description of the algorithm

Let n_k be *splitting factors* (in our example $n_k = 2$) and r be the total number of paths that reached set B. We put

$$\widehat{p} := \frac{r}{\prod_{k=0}^{m-1} n_k}$$

$\widehat{p}$ is an unbiased estimator of p for ANY choice of intermediate sets.

For a good choice of intermediate sets, the total computational cost of the estimation is proportional to $(\log p)^2$ instead of p^{-1}!

PART II

Rare Event Simulation for the Stationary Distribution of a Markov Chain

Stationary Markov Chains

Let $X = (X_n)_{n \in \mathbb{N}}$ be an $\mathbb{R}^d$-valued, time-discrete Markov chain with *stationary (invariant)* measure μ, that is, as $n \to \infty$,

$$X_n \rightsquigarrow X_\infty \sim \mu.$$

Context: numerical solutions to SDEs:

$$\mathrm{d}X_t = f(X_t)\mathrm{d}t + g(X_t)\mathrm{d}W_t$$

and

$$X_{n+1} = X_n + f(X_n)h + g(X_n)\sqrt{h}\,\Delta W_n$$

We want to estimate $\mu(B)$ when $\mu(B) \ll 1$. From Ergodic Theorem, for any set B:

$$\mu(B) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}\{X_n \in B\}.$$

Recurrent Structure of a Markov Chain

Consider the following decomposition of a Markov chain:

- Choose a set $A \subset \mathbb{R}^d$, with $\mu(A) \in (0, 1)$.
- Let S_k be consecutive *inwards crossings* of set A, with $S_{-1} = 0$,

$$S_k := \inf\{n > S_{k-1} : X_{n-1} \notin A, X_n \in A\}.$$

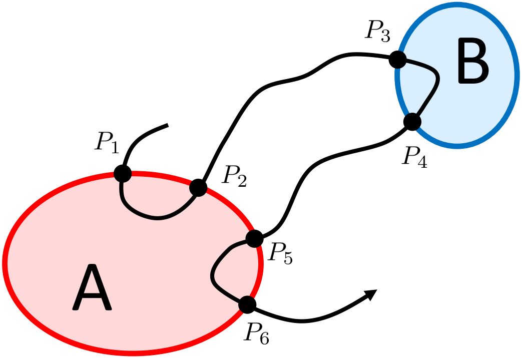
- Let $\mathcal{C}_k$ be a path within kth cycle, i.e.

$$\mathcal{C}_k := \big(X_n : S_{k-1} \leq n < S_k\big).$$

- We distinguish the cycle *length* and *origin*

$$L_k := S_k - S_{k-1}, \quad X_k^A := X_{S_{k-1}}$$

Note: Assuming the chain is 'sufficiently nice', we have $\mathbb{E}L_k < \infty$.



The cycle begins at P_1 and ends at P_5.

Recurrent Structure of a Markov Chain

In the stationary regime, the cycles $\mathcal{C}_1, \mathcal{C}_2, \ldots$ are **identically distributed**.

Define the *time spent in set B* within a cycle:

$$R_k := \sum_{n=S_{k-1}}^{S_k - 1} \mathbb{1}\{X_n \in B\}.$$

The quantity of interest $p = \mu(B)$ can be expressed as:

$$p = \frac{\mathbb{E}R_1}{\mathbb{E}L_1}.$$

- Estimation of $\mathbb{E}L_1$ is easy (Monte Carlo).

Estimation of $\mathbb{E}R_1$

Notice that

$$T_B := \mathbb{E}R_1 = \mathbb{P}(\tau_B < \tau_A^{\text{in}}) \cdot \mathbb{E}(R_1 \mid R_1 > 0)$$

This fits in the framework of MLS with an extra stage with splitting factor n_m!

$$\widehat{T}_B := \frac{r_m}{\prod_{k=0}^{m-1} n_k} \cdot \frac{\sum_{j=1}^{n_m r_m} \widehat{R}_+^{(j)}}{n_m r_m}$$

$\widehat{T}_B$ is an unbiased estimator for $\mathbb{E}R_1$!

For a good choice of intermediate sets and set A, the computational cost is proportional to $(\log p)^2$!

Assumptions leading to optimality

The study of efficiency of the algorithm is intractable for a general choice of intermediate sets. Recall that

$$\tau_k = \inf\{n > 0 : X_n \in B_k\}, \quad D_k := \{\tau_k < \tau_A\}$$

We assume the following:

(I) for all $k \in \{1, \dots, m-1\}$, for all X_{τ_k},

$$\mathbb{P}(D_{k+1} \mid D_k, X_{\tau_k}) \equiv \mathbb{P}(D_{k+1} \mid D_k)$$

(II) for all X_1^A,

$$\mathbb{P}(\tau_B < \tau_A \mid X_1^A) \equiv \mathbb{P}(\tau_B < \tau_A)$$

(III) for all cycle origins X_{τ_B},

$$\left(R_1 \mid R_1 > 0, X_{\tau_B}\right) \overset{d}{=} (R_1 \mid R_1 > 0) =: R_+,$$

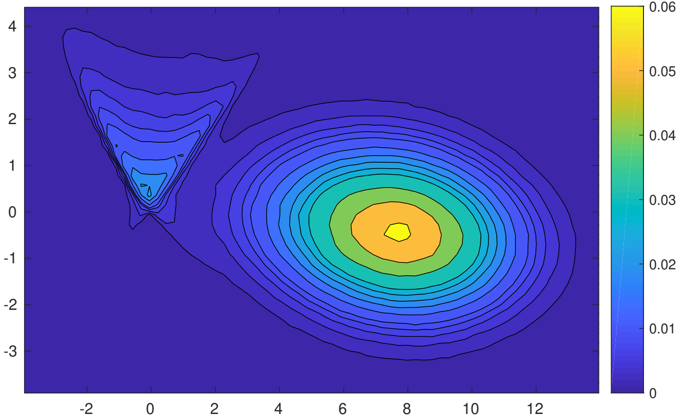
Numerical Example: Franzke (2012) Model

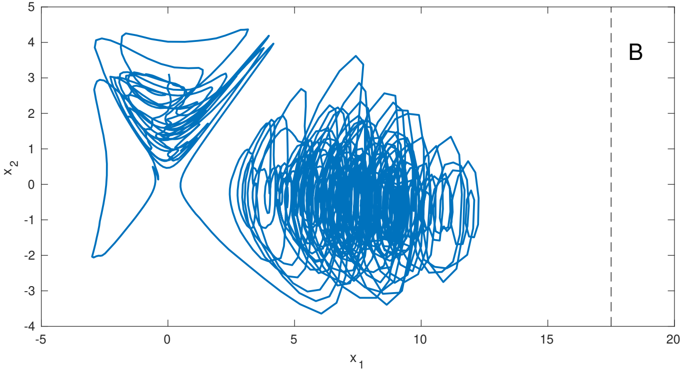
Franzke (2012) Model

$$\begin{aligned}
\mathrm{d}x_1 &= \mu\big(-x_2(L_{12} + a_1 x_1 + a_2 x_2) + d_1 x_1 + F_1 \\
&\qquad + L_{13} y_1 + B^1_{123} x_2 y_1 + (B^2_{131} + B^2_{113}) x_1 y_1 \big)\mathrm{d}t \\
\mathrm{d}x_2 &= \mu\big(+x_1(L_{21} + a_1 x_1 + a_2 x_2) + d_2 x_2 + F_2 \\
&\qquad + L_{24} y_2 + B^1_{213} x_1 y_1 + (B^3_{242} + B^3_{224}) x_2 y_2 \big)\mathrm{d}t \\
\mathrm{d}y_1 &= \mu\big(-L_{13} x_1 + B^1_{312} x_1 x_2 + B^2_{311} x_1^2 + F_3 - \tfrac{\gamma_1}{\varepsilon} y_1 \big)\mathrm{d}t + \tfrac{\sigma_1}{\sqrt{\varepsilon}}\mathrm{d}W_1 \\
\mathrm{d}y_2 &= \mu\big(-L_{24} x_2 + B^3_{422} x_2 x_2 + F_4 - \tfrac{\gamma_2}{\varepsilon} y_2 \big)\mathrm{d}t + \tfrac{\sigma_2}{\sqrt{\varepsilon}}\mathrm{d}W_2
\end{aligned}$$

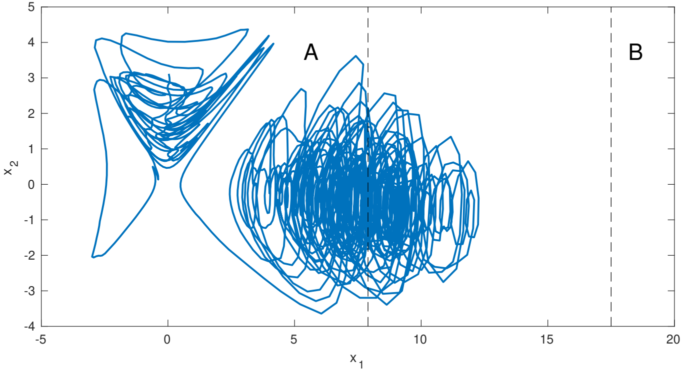
$$p = \lim_{n\to\infty} \mathbb{P}(x_1 > u) = \mu(x_1 > u)$$

Franzke (2011) Model





5
4
3
2
1
0
-1
-2
-3
-4
x 2
-5
0
5
10
15
20
x 1
B



A
B
x 1
x 2
-5
0
5
10
15
20
-4
-3
-2
-1
0
1
2
3
4
5

Results

u	14	15	16	17.5	18.5
$\widehat{p}$	$1.08 \cdot 10^{-3}$	$1.99 \cdot 10^{-4}$	$3.00 \cdot 10^{-5}$	$1.14 \cdot 10^{-6}$	$9.78 \cdot 10^{-8}$
$\mathrm{Eff}(\widehat{p})$	1.9	8.6	32.1	269.9	1521.8

Table: RMS algorithm applied to Franzke (2012) model. Parameters: $A = \{x_1 \leq 7.9\}$, $B = \{x_1 > u\}$. Importance function is $H(\mathbf{x}) = \frac{x_1}{u}$. The relative errors are below 1%.

This approach is orders of magnitude faster than Monte Carlo!

Conclusions

- We presented an algorithm for the estimation of rare events associated with the stationary distribution of a Markov chain.
- Implementation of the algorithm does not require any knowledge of the system under study – **it can be applied to 'black-box' models.**
- Open question: good choice of the recurrency set A.

References

1. Bisewski, K., Crommelin, D. and Mandjes, M., 2019. Rare event simulation for steady-state probabilities via recurrency cycles. Chaos: An Interdisciplinary Journal of Nonlinear Science, 29(3), p.033131.
2. Franzke, C., 2012. Predictability of extreme events in a nonlinear stochastic-dynamical model. Physical Review E, 85(3), p.031134.

Finishing remarks

Choice of the recurrency set A:

(i) Recall that A should be such that for all cycle origins X_1^A:

$$\mathbb{P}(\tau_B < \tau_A \,|\, X_1^A) \equiv \mathbb{P}(\tau_B < \tau_A)$$

(ii) At the same time A should be such that $\mathbb{E}L_1$ is not 'too large', so A should not be 'too small' ($\mu(A) \approx 0$) nor 'too big' ($\mu(A) \approx 1$).

Numerical implementation:

1. Estimate $\mathbb{E}L_1$ using Monte Carlo method. Store the locations of the cycle origins in the set $\mathcal{S} := \{X_{S_0}, X_{S_1}, \ldots\}$.
2. Estimate T_B using Multilevel Splitting. Bootstrap cycle origins from the set $\mathcal{S}$.

Optimal Parameters

We aim to minimize the **computational time** of the algorithm under the constraint $\mathrm{RE}^2(\widehat{T}_B) = \frac{\mathbb{V}\mathrm{ar}\, T_B}{(\mathbb{E}\,\widehat{T}_B)^2} < q^2$ for a chosen $q > 0$.

$$
\begin{aligned}
m &= c\,|\log p|,\\
p_k &= \frac{2c-1}{2c} \approx \frac{1}{5}, \quad k \in \{1, \dots, m\},\\
n_k &= 1/p_{k+1} \approx 5, \quad k \in \{1, \dots, m-1\},\\
n_m &= \mathrm{RE}(R_+) \cdot \frac{2c}{\sqrt{2c-1}},\\
n_0 &= \frac{1}{q\sqrt{2c-1}} \cdot \left(\frac{c\,|\log p|}{\sqrt{2c-1}} + \mathrm{RE}(R_+) \right),\\
W(\widehat{T}_B) &\propto \frac{1}{q} \left(\frac{c\,|\log p|}{\sqrt{2c-1}} + \mathrm{RE}(R_+) \right)^2.
\end{aligned}
$$

with $c \approx 0.6275$ solving $\exp(1/c) = 2c/(2c-1)$.

How do we choose the importance function?

Let $H : \mathbb{R}^d \longrightarrow [0,1]$ be the **importance function** and for *levels* $0 = l_0 < l_1 < \ldots < l_m = 1$ we put

$$B_k := \{x \in \mathbb{R}^d : H(x) \geq l_k\}.$$

Ideal importance function H should satisfy:

$$\boxed{H(x) \geq H(y) \implies \mathbb{P}_x(\tau_B < \tau_A) \geq \mathbb{P}_y(\tau_B < \tau_A).}$$

In particular

$$H(x) := \mathbb{P}(\tau_B < \tau_A \,|\, X_0 = x)$$

satisfies the above and so is

$$H_g(x) := g(H(x))$$

for any increasing function g.