Project 7: Stochastic modeling of pollutant transport in aquifers

Matteo Calafà

École Polytechnique Fédérale de Lausanne (EPFL)

Course of Stochastic Simulation

1 Feb 2022

EPFL

Outlines

Crude Monte Carlo

The pollutant trajectory can be discretized with a Euler-Maruyama scheme:

$$\begin{cases} X_{k+1} = X_k + u_1(X_k, Y_k)\Delta t + \sigma\sqrt{\Delta t} Z_k & Z_k \sim \mathcal{N}(0,1) \\ Y_{k+1} = Y_k + u_2(X_k, Y_k)\Delta t + \sigma\sqrt{\Delta t} Z'_k & Z'_k \sim \mathcal{N}(0,1) \end{cases}$$

A solution to estimate p is now straightforward: simulate many discretized trajectories and do a simple Monte Carlo estimation. However:

- Slow convergence rate
- Not suitable for rare events.

Feynman-Kac equivalent problem

This is a deterministic equivalent problem that can be solved with standard numerical methods. Its homogeneous weak formulation to be used in Finite Element libraries as `Fenics` follows:

$$V_h := \{f \in C^0(D) : \quad f|_{\partial D} = 0, \quad f|_{t_h} \in \mathbb{P}^1(t_h)\, \forall t_h \in \mathcal{T}_h\}$$

$$\text{Find } \{\varphi_n\}_{n=0\ldots N} \subset V_h \text{ s.t. } \varphi_0 = 0 \text{ and } \forall n = 1 \ldots N-1 :$$

$$\int_D \varphi_n v - \int_D \varphi_{n-1} v - \Delta t \int_D (\mathbf{u} \cdot \nabla)\varphi_n\, v + \frac{\Delta t}{2}\sigma^2 \int_D \nabla\varphi_n \cdot \nabla v = 0$$

where first order implicit Euler has been applied to the temporal step and first order polynomials for the F.E. space.

Importance Sampling motivation

This is a variance reduction technique that is well-suited for Markov chains with stopping times, indeed:

Proposition

Suppose (p_0, P), (q_0, Q) two continuous-space discrete-time Markov chains and p, q their density functions, then:

$$\mathbb{E}_p[\psi_\tau(X_0, \dots, X_\tau)] = \mathbb{E}_q[\psi_\tau(X_0, \dots, X_\tau) w(X_0, \dots, X_\tau)]$$

$$\text{where} \quad w(X_0, \dots, X_\tau) = \frac{p_0(X_0)}{q_0(X_0)} \prod_{j=1}^{\tau} \frac{p(X_{j-1}, X_j)}{q(X_{j-1}, X_j)}$$

IS for the pollutant problem

Our choice for the *importance distribution* is to shift the original normal kernel. The new Markov kernel is then:

$$\begin{cases} X_{k+1} = X_k + (u_1(X_k, Y_k) + c_x)\Delta t + \sigma\sqrt{\Delta t} Z_k \\ Y_{k+1} = Y_k + (u_2(X_k, Y_k) + c_y)\Delta_t + \sigma\sqrt{\Delta t} Z'_k \end{cases} \qquad c_x, c_y \in \mathbb{R}$$

From this choice, the correction factor w can consequently be computed. Moreover, c_x, c_y need to be properly set in order to effectively decrease the variance. An empirical study has been reproduced and it is shown later.

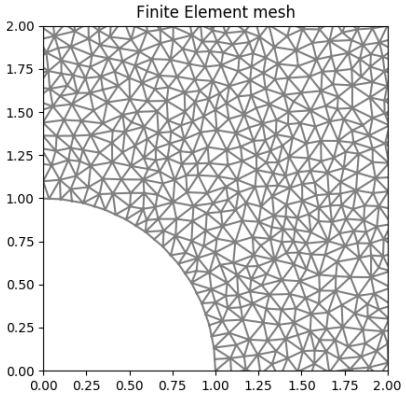
Splitting Method for rare events

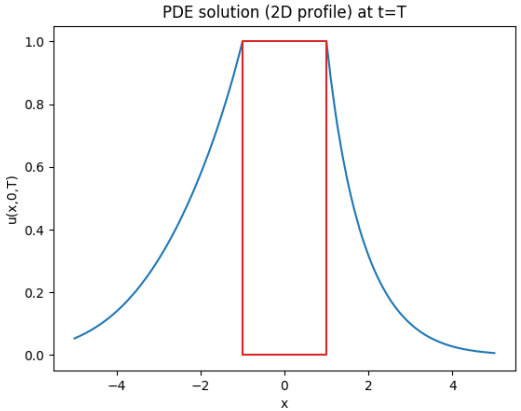
Consider a trajectory that starts far from the well. The idea of the Splitting Method [1] is to split the total trajectory into different domains. The main implementation choices:

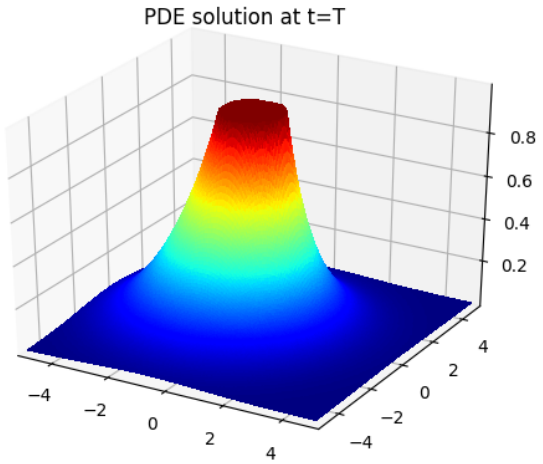
- The subdomains are concentric balls with constant radius increments (reasonable from the spherical domain shape).
- The number of simulated trajectories is constant for every subdomain (*Fixed Effort Splitting*, [2]).
- The starting points of a subdomain are sampled randomly (and necessarily with repetition) from the successful ending points of the external subdomain.
- The number of layers is taken from the optimal choice $L = -\ln(p_{true})/2 \approx 7$.

Results

Finite Element solution







Finite Element probability estimations

The regularity of the solution, the grid independence and the fulfilment of the boundary conditions have been checked. In addition, we performed some simulations to check the independence of the probability estimations from:

- The mesh refinement
- The number of temporal steps
- The domain size

These gave us excellent results.
With 719'767 nodes, 800 steps, 40R domain radius, the reference solutions for the entrance probabilities are:

(1.2, 1.1)	(3.0, 4.0)	(2.5, 2.5)	(7.0, 7.0)
50.8%	0.98%	6.26%	$5 \cdot 10^{-7}$

Crude Monte Carlo results

	$5 \cdot 10^2$		$5 \cdot 10^3$		$5 \cdot 10^4$	
$\mathbf{X}_0$	Prediction	C.I.	Prediction	C.I.	Prediction	C.I.
(1.2,1.1)	0.5100	0.0576	0.5052	0.0182	0.5041	0.0058
(3.0,4.0)	0.0064	0.0089	0.0104	0.0037	0.0097	0.0011
(2.5,2.5)	0.0612	0.0273	0.0654	0.0090	0.0064	0.0028

Table: Varying the number of iterations (time-step $= 10^{-4}$)

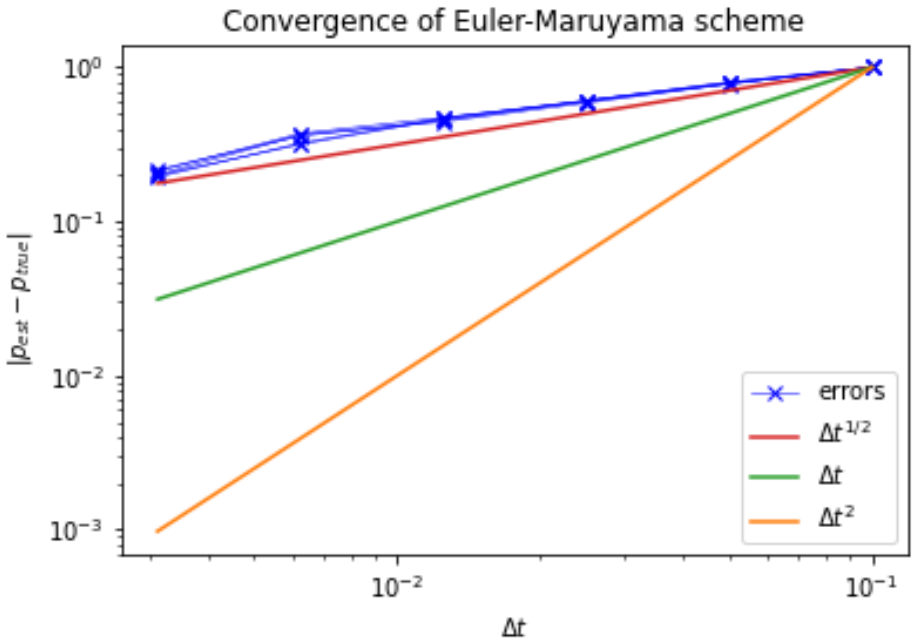
	10^{-2}		10^{-3}		10^{-4}	
$\mathbf{X}_0$	Prediction	C.I.	Prediction	C.I.	Prediction	C.I.
(1.2,1.1)	0.44342	0.00571	0.48638	0.00576	0.50412	0.00575
(3.0,4.0)	0.00798	0.00100	0.00844	0.00110	0.00966	0.00105
(2.5,2.5)	0.05412	0.00257	0.05896	0.00270	0.06248	0.00274

Table: Varying the time discretization ($5 \cdot 10^4$ Monte Carlo iterations)

- C.I. decreases with $\sqrt{10}$ rate in the first, constant in the other.
- Results converge to the reference ones as parameters improve.

Euler discretization error study

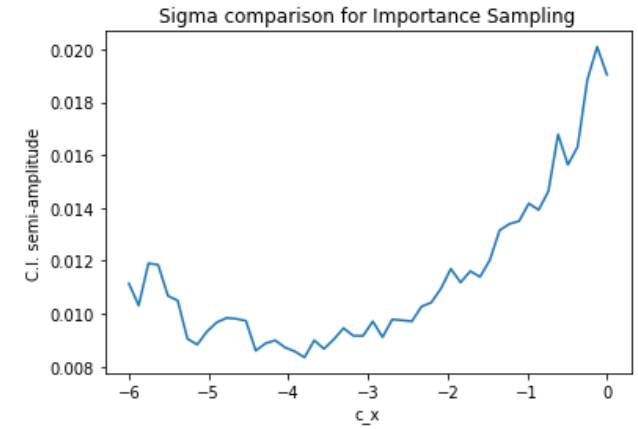
Idea: to reduce randomicity, very coarse trajectories with instead many Monte Carlo iterations and shared Brownian increments.

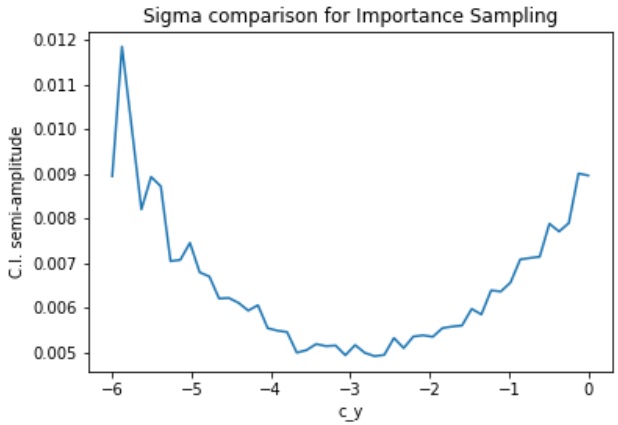


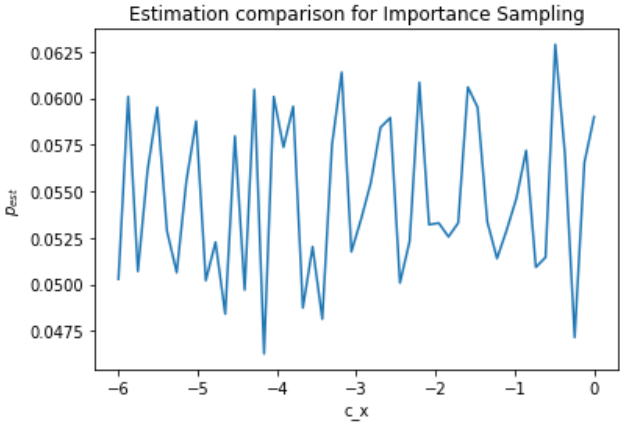
From the 5 convergence tests, it is clear that the discretization error has a 1/2 order.

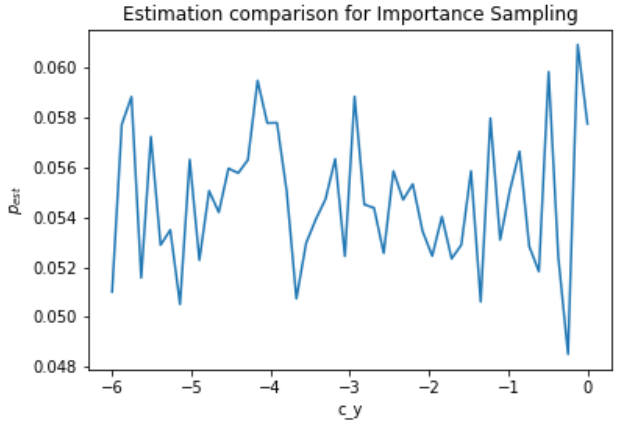
Importance Sampling parameters selection

Idea: coordinate descent, fix one parameter and optimize the other.









(a) c_x varies, $c_y = 0$

(b) c_y varies, $c_x = -4$

Importance Sampling variance reduction

From the previous plots, it is clear that $c_x = -4, c_y = -3$ is the best set of parameters. Moreover, the parameters choice is proven not to influence the estimations.

With this choice, results obtained with the same discretization values as before turned out to be consistent but, this time, with a C.I. reduced on average 4 times.

Splitting method results

$\mathbf{x}_0$	Run 1	Run 2	Run 3	Run 4	Mean
(7.0,7.0)	4.39580e-07	4.76504e-07	3.90226e-07	5.67980e-07	4.68573e-07

- 10^4 M.C. iterations per level, time-step $= 5 \cdot 10^{-4}$, 7 levels.
- These values are coherent with the $5 \cdot 10^{-7}$ F.E. solution.
- Since the number of layers is 7, the averaged conditional probability is ≈ 0.10, a value that can reasonably be computed also with standard Monte Carlo. The full result, instead, is so small that the probability that Monte Carlo returns the exact zero is approximately 98%.

References

[1] Marnix Joseph Johann Garvels. "The splitting method in rare event simulation". PhD thesis. University of Twente, 2000.

[2] P. Kroese Dirk, Thomas Taimre, and I. Botev Zdravko. *Handbook of Monte Carlo Methods*. New York: John Wiley & Sons, 2011. ISBN: 9781118014967. DOI: https://doi.org/10.1002/9781118014967.