

VARIANCE REDUCTION FOR CHEMICAL REACTION NETWORKS WITH ARRAY-RQMC

Florian Puchhammer

Université de Montréal, Canada

MCM 2019, Sydney

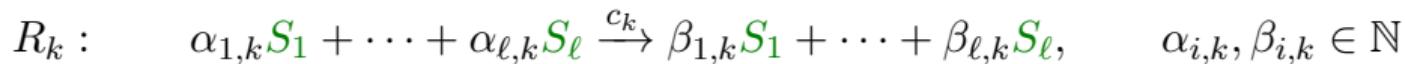
July 12, 2019



Joint work with **Amal Ben Abdellah** and **Pierre L'Ecuyer**

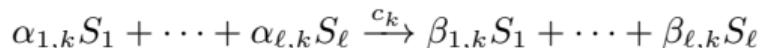
Chemical reaction networks

Consider chemical species S_1, S_2, \dots, S_ℓ that react via reactions R_1, R_2, \dots, R_K with reaction rates c_1, c_2, \dots, c_K .



Goal: Simulate system over time $t \in [0, T]$ and observe (a function g of) copy numbers $\mathbf{X}_t = (X_{1,t}, X_{2,t}, \dots, X_{\ell,t})$ of the species.

The τ -leap algorithm



Simulate with fixed step τ -leap algorithm, Gillespie '01. Summarizes several reactions into one time step of length $\tau \rightsquigarrow$ faster, but introduces bias (negative copy numbers).

- Poisson variables p_k model how often reaction R_k fires in $[t, t + \tau)$.
- Mean parameters of p_k are $a_k(\mathbf{X}_t)\tau$, where a_k are the propensity functions.
- a_k 's are polynomials in \mathbf{X}_t of degree $\alpha_{1,k} + \alpha_{2,k} + \dots + \alpha_{\ell,k}$.

In each step, generate p_1, \dots, p_K and update copy numbers through

$$X_{i,t+\tau} = X_{i,t} + \sum_{k=1}^K p_k \cdot (\beta_{i,k} - \alpha_{i,k}).$$

This is a discrete time Markov chain (DTMC), with ℓ -dimensional state space using K random variates to advance the chain by one step.

Simulation

Goal: Estimate

$$\mu = \mathbb{E}[g(\mathbf{X}_T)].$$

MC: Generate n points $\mathbf{U}_i \sim \mathcal{U}(0, 1)^{K \cdot T / \tau}$.

- Each \mathbf{U}_i simulates one chain $\rightsquigarrow \mathbf{X}_{i,T}$.
- Estimate

$$\mu \approx \frac{1}{n} \sum_{i=0}^{n-1} g(\mathbf{X}_{i,T}).$$

RQMC: Generate n points $\mathbf{U}_i \in (0, 1)^{K \cdot T / \tau}$ such that:

- Each \mathbf{U}_i is **uniformly distributed** in $(0, 1)^{K \cdot T / \tau}$.
- The point set $\{\mathbf{U}_0, \mathbf{U}_1, \dots, \mathbf{U}_{n-1}\}$ has **low discrepancy**.
- Then, estimate as above.

Previous work

Beentjes, Baker '18: Traditional RQMC for τ -leaping. Found that, using n RQMC points, the convergence rate is not better than for MC, n^{-1} , even when $\ell, K, T/\tau$ were not too big.

Possible problems:

- Discontinuities because of integer values.
- High dimensional point sets needed, $\text{dim} = K \cdot T/\tau$ (actually effective dimension important).

Possible solution: array-RQMC.

- Can work well even when discontinuities appear.
- Reduces the dimensionality of the points to $\ell + K$.

Array-RQMC

Simulation of n chains with n RQMC points.

Traditional approach: One point for the trajectory of one chain. Simulates chains **sequentially** \rightsquigarrow high dimension $K \cdot T/\tau$.

Array-RQMC: One point to advance one chain by one step. Advances all n chains in **parallel** at each step \rightsquigarrow lower dimension $\ell + K$.

- **Idea:** the distance between empirical and theoretical distribution of states is small at each step.
- achieved by **matching** the states to **first ℓ coordinates** of point set via (multivariate) sort after each step.
- **empirical mean** of a n evaluations of function g at the final states is **unbiased** estimator for the mean of $g(\mathbf{X}_T)$.
- **empirical variance** of m independent repetitions is **unbiased** estimator for the variance of the mean estimator.

Multivariate sorting algorithms

If $\ell = 1$ sort chains by size. If $\ell > 1$:

Split sort: Split states into 2 packets by size of first coordinate. Split each packet into 2 by size of second coordinate. . .

Batch sort: Split states into n_1 packets by size of first coordinate. Split each packet into n_2 by size of second coordinate. . .

Hilbert curve sort: Map states to $[0, 1]^\ell$. Sort the states in the order they are visited by (discrete version of) hilbert curve. Here, we use **sample data** to estimate mean and variance, normalize each state, and use the **CDF of standard normal distribution** to map states to $[0, 1]^\ell$.

Customized sorts: Ideally, find **importance function** $h_j : \mathbb{R}^\ell \rightarrow \mathbb{R}$ for step j , then sort by size of $h_j(\mathbf{X}_{j\tau})$. Might increase efficiency!

Experimental setting

In every experiment, we observe copy number of one specific species.

Point sets:

- **MC:** Plain monte carlo
- **Lat+Baker:** Lattice rule with random shift mod 1 and baker transformation.
- **Sob+LMS:** Sobol' points with a left matrix scramble and a digital shift.
- **Sob+NUS:** Sobol' points with Owen's nested uniform scramble.

Observed statistics: We replicate each experiment $m = 100$ times independently.

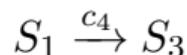
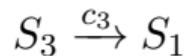
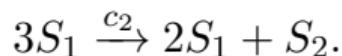
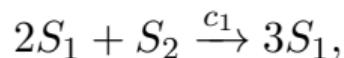
- **VRF20:** Variance reduction factor in comparison to MC for $n = 2^{20}$ points.
- $\hat{\beta}$: Variance convergence rate $n^{-\hat{\beta}}$ estimated by regression $n = 2^{13}, 2^{14}, \dots, 2^{20}$.

Goal: Show that

- array-RQMC can beat MC-variance rate $\hat{\beta} > 1$,
- **customized sorts** can improve your results (**proof of concept**).

The Schlögl system

Model with 3 species S_1, S_2, S_3 and $K = 4$ reactions. We observe S_1 .



Note: $X_{1,t} + X_{2,t} + X_{3,t} = \text{const.}$, therefore $\ell = 2$. Expect standard sorts to work well.

Simulation: $\mathbf{X}_0 = \{250, 10^5, 2 \cdot 10^5\}$, $c_1 = 3 \cdot 10^{-7}$, $c_2 = 10^{-4}$, $c_3 = 10^{-3}$, $c_4 = 3.5$, $T = 4$, $\tau = T/15$.

Array-RQMC reduces points' dimension from $KT/\tau = 60$ to $\ell + K = 6$ or 5.

Customized sorts revisited

Observe that $\ell = 2$, so we can plot final copy number of S_1 vs. current state at time $j\tau \rightsquigarrow$ importance function h_j .

- Simulate 2^{20} sample chains with MC.
- Fit polynomial to this data.
- To reduce terms: take **propensity functions** and current copy number of observed species.

Here (N_0 is total number of molecules):

$$a_1(x, y) = \frac{1}{2}c_1x(x-1)y,$$

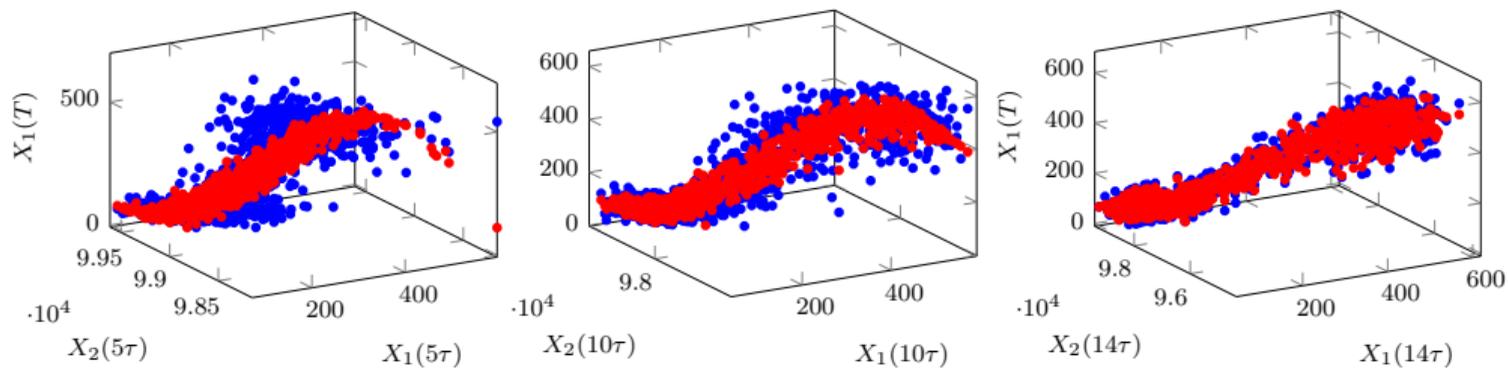
$$a_2(x, y) = \frac{1}{6}c_2x(x-1)(x-2),$$

$$a_3(x, y) = c_3(N_0 - x - y),$$

$$a_4(x, y) = c_4x.$$

Take $h_j(x, y) = a_{j,0} + a_{j,1}x + a_{j,2}y + a_{j,3}x^2 + a_{j,4}xy + a_{j,5}x^3 + a_{j,6}x^2y$.

The **sample data** at step $j\tau$ and **fitted** h_j , $j = 5, 10, 14$.



For **customized sorts** we can follow two strategies:

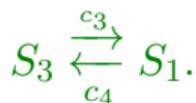
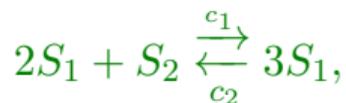
Multi step: Fit h_j in every step.

- Expected to be more representative.

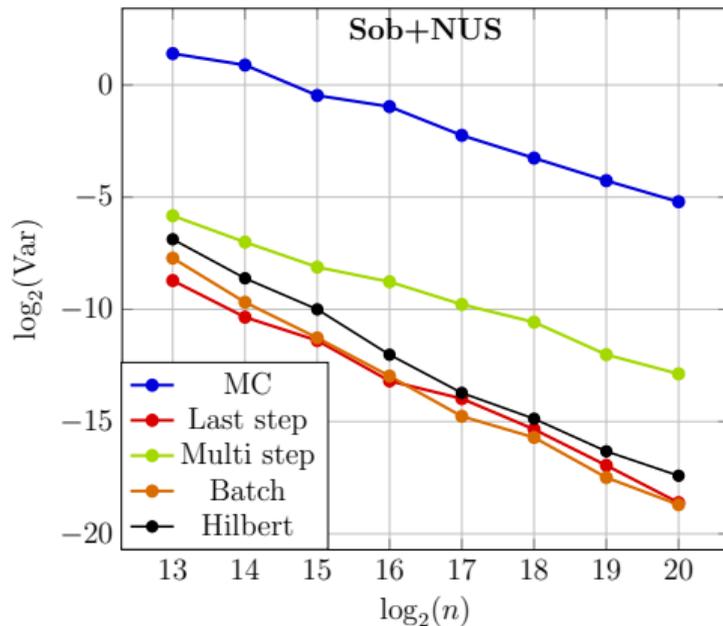
Last step: Fit $h_j = h$ to last step only. Use this h in every step.

- Maybe less impact of randomness on data.

Schlögl system – results

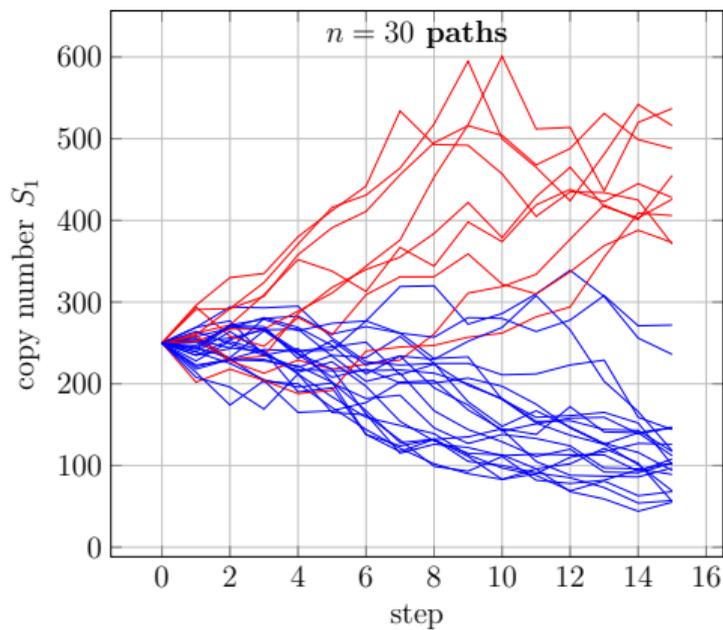
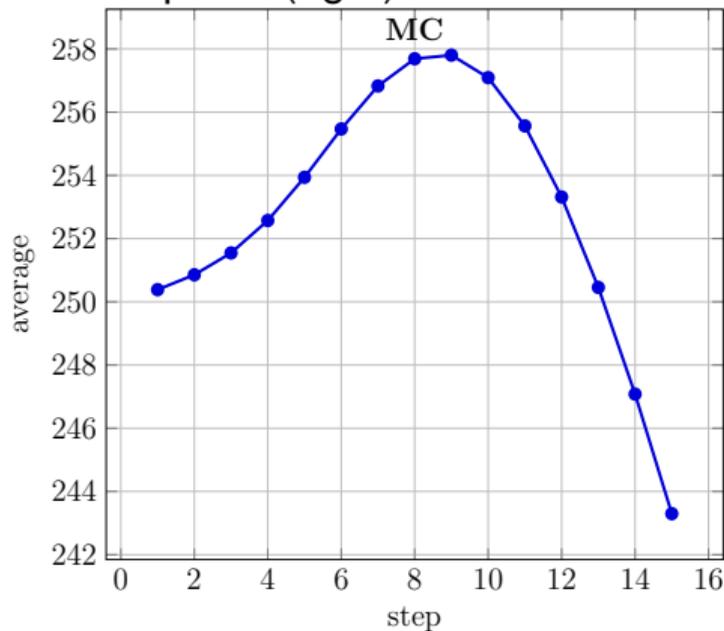


Sort	Sample	$\hat{\beta}$	VRF20
	MC	1.00	1
Last step	Lat+Baker	1.30	4894
	Sob+LMS	1.36	7000
	Sob+NUS	1.37	10481
Multi step	Lat+Baker	1.01	178
	Sob+LMS	1.07	206
	Sob+NUS	0.99	196
Batch	Lat+Baker	1.39	4421
	Sob+LMS	1.48	9309
	Sob+NUS	1.56	11150
Hilbert curve	Lattice+Baker	1.68	2493
	Sobol+LMS	1.51	4464
	Sobol+NUS	1.53	4562



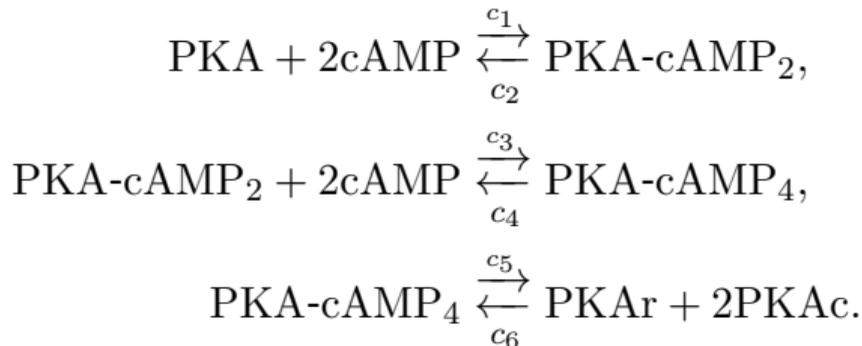
Problems with multi step sort

Look at mean copy numbers per step with MC (left) and copy numbers of S_1 for $n = 30$ paths (right).



cAMP aktivation of PKA model

More challenging example: $\ell = 6$ species, $K = 6$ reactions (Koh, Blackwell ('12); Strehl, Ilie ('15)):



Initial values: PKA: $33 \cdot 10^3$, cAMP: $33.03 \cdot 10^3$, others: $1.1 \cdot 10^3$.

Parameters: $c_1 = 8.696 \cdot 10^{-5}$, $c_2 = 0.02$, $c_3 = 1.154 \cdot 10^{-4}$, $c_4 = 0.02$, $c_5 = 0.016$ and $c_6 = 0.0017$; $T = 5 \cdot 10^{-5}$, $\tau = T/15$.

Array-RQMC reduces points' dimension from $KT/\tau = 90$ to $\ell + K = 12$ or 7.

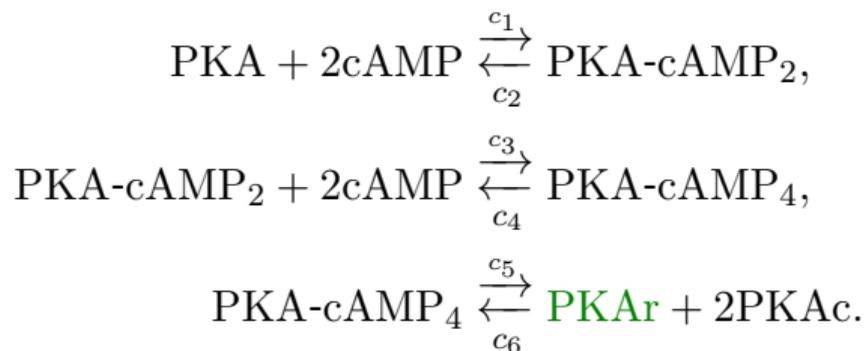
Begin with PKA and use same heuristics (sample paths, propensity functions, and current copy number of PKA) for last step and multi step.

Sort	Sample	$\hat{\beta}$	VRF20
	MC	1.00	1
Last step	Lat+Baker	1.05	2593
	Sob+LMS	1.05	3334
	Sob+NUS	1.04	3126
Multi step	Lat+Baker	1.03	2492
	Sob+LMS	1.04	3307
	Sob+NUS	1.01	3099
Batch	Lat+Baker	1.05	2575
	Sob+LMS	1.06	3611
	Sob+NUS	1.01	2735
Hilbert curve	Lattice+Baker	1.00	2327
	Sobol+LMS	1.03	2920
	Sobol+NUS	1.02	2847

- Variance rate aligns with MC rate.
- Best result for Batch with Sob+LMS. Last step makes up for it with computation time.
- Reducing noise in sample paths improved Last step up to VRF20=4100 (Sob+NUS).
- Select “optimal” sequence of last step and multi step sorts w.r.t. variance increment at each step improved up to VRF20=4300 (Sob+LMS).

Dimensionality is NOT the key

Same model, but now we observe $PKAr$:

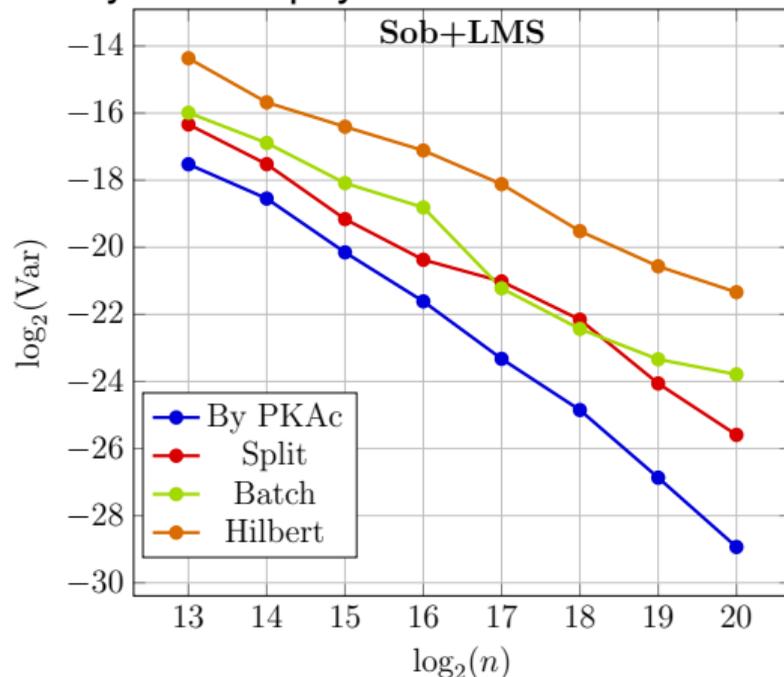


- Analyzed the data and found indicator, that only $PKAc$ is important for development of $PKAr$.
- Tried to simply sort by $PKAc$.

Results for PKAr

Exploiting additional information to customize your sort pays off!

Sort	Sample	$\hat{\beta}$	VRF20
	MC	1.03	1
By PKAc	Lattice+Baker	1.59	5187
	Sobol+LMS	1.63	19841
	Sobol+NUS	1.61	16876
Split	Lattice+Baker	1.06	1021
	Sobol+LMS	1.28	1958
	Sobol+NUS	1.26	1756
Batch	Lattice+Baker	1.16	803
	Sobol+LMS	1.22	561
	Sobol+NUS	1.25	744
Hilbert curve	Lattice+Baker	1.01	103
	Sobol+LMS	1.00	103
	Sobol+NUS	1.04	106



Conclusion

- Introduced chemical reaction networks.
- Presented τ -leap algorithm, which allows to simulate via DTMCs.
- With traditional RQMC, no improvement in convergence of variance (Beentjes, Baker ('18)).
- We investigate same problem with array RQMC, which relies on sorting strategies for the states.
- In empirical studies we showed that standard multivariate sorts work very well.
- When states have higher dimension, customized sorts can increase efficiency.
- Using additional info/heuristics to customize own sort can improve results tremendously!

References – chemical reaction networks.

- C. H. L. Beentjes, Baker, Ruth E., *Quasi-Monte Carlo Methods Applied to Tau-Leaping in Stochastic Biological Systems*. Bulletin of Mathematical Biology, 2018.
- Y. Cao, D. T. Gillespie, L. R. Petzold *Avoiding negative populations in explicit Poisson tau-leaping*. Journal of Chemical Physics 123. 054104, 2005.
- Y. Cao, D. T. Gillespie, L. R. Petzold *Efficient step size selection for the tau-leaping simulation method*. Journal of Chemical Physics 124. 044109, 2006.
- D. T. Gillespie., *Approximate accelerated stochastic simulation of chemically reacting systems*. The Journal of Chemical Physics 115. 1716–1733, 2001.
- D. T. Gillespie., *Exact stochastic simulation of coupled chemical reactions*. The Journal of Physical Chemistry 81. 2340–2361, 1977.
- W. Koh, K. T. Blackwell, *Improved spatial direct method with gradient-based diffusion to retain full diffusive fluctuations*. The Journal of Chemical Physics 137. 154111, 2015.
- J.M.A. Padgett, S. Ilie, *An adaptive tau-leaping method for stochastic simulations of reaction-diffusion systems*. AIP Advances 6. 035217, 2016.
- R. Strehl, S. Ilie, *Hybrid stochastic simulation of reaction-diffusion systems with slow and fast dynamics*. The Journal of Chemical Physics 143. 234108, 2015.

References – array-RQMC.

- A. Ben Abdellah, P. L'Ecuyer, F. Puchhammer *Array-RQMC for option pricing under stochastic volatility models*. submitted, 2019 (available on arXiv).
- M. Gerber, N. Chopin, *Sequential quasi-Monte Carlo*. Journal of the Royal Statistical Society, Series B, 77. 509–579, 2015.
- P. L'Ecuyer, V. Demers, B. Tuffin, *Rare-Events, Splitting, and Quasi-Monte Carlo*. ACM Transactions on Modeling and Computer Simulation 17. Article 9, 2007.
- P. L'Ecuyer, C. Lécot, A. L'Archevêque-Gaudet, *On Array-RQMC for Markov Chains: Mapping Alternatives and Convergence Rates*. Monte Carlo and Quasi-Monte Carlo Methods 2008. 485–500, 2009.
- P. L'Ecuyer, C. Lécot, B. Tuffin, *A randomized quasi-Monte Carlo simulation method for Markov chains*. Operations Research 56. 958–975, 2008.
- P. L'Ecuyer, D. Munger, C. Lécot, B. Tuffin, *Sorting methods and convergence rates for array-rqmc: Some empirical comparisons*. Mathematics and Computers in Simulation. 2017.