Particle allocation for steady state sampling

David Aristoff (joint work with Daniel M. Zuckerman); thanks to Mathias, Tony, Gabriel, Gideon

Colorado State University

September 2018

Goal: variance reduction in steady-state computations

Motivation: computing MFPT by Hill relation

$$\text{Mean first passage time to } F = \frac{1}{\text{steady-state flux into } F}$$

More generally: Estimate $\int f\,d\mu$ where $\mu =$ stationary distribution of a (usually nonreversible) Markov kernel K

A weighted ensemble

$$\{\text{parent particles, weights}\} \xrightarrow{\text{selection}} \{\text{children particles, new weights}\} \xrightarrow{\text{mutation}} \{\text{new parent particles, weights}\}$$

In a **selection** step,

the parent particles are resampled to get children

In a **mutation** step,

all the children evolve one step via Markov kernel K

Stratification: space divided into bins $r \in \mathcal{R}$

Allocation strategy: choose # of children in each bin

$$N_t(r) := \#\text{children in bin } r \text{ at time/generation } t$$

Think of bins as rectangles/Voronoi cells in low-D space of reaction coordinates (as in WE, EM, NEUS, ...)

How to choose N_t? And bins? Resampling times? etc...

Optimal allocation:

$$N_t(r) \approx \frac{N\omega_t(r)v_t(r)}{\sum_r \omega_t(r)v_t(r)}$$

where

$$\begin{aligned} N &= \text{fixed total \# of particles} \\ \omega_t(r) &= \text{weight in bin } r \text{ at time } t \\ v_t &= \text{selection value function} \end{aligned}$$

Can derive this formula, and v_t, from first principles

In practice: Use bin-to-bin transitions to learn v_t

$Q_t(r, s) \approx Pr(\text{a child goes to bin } s | \text{its parent is from bin } r)$

Q_t **updated adaptively in steady-state simulations**

Optimal strategy:

π_t = stationary vector of Q_t : $\quad \pi_t^T Q_t = \pi_t^T, \; \sum_{r \in \mathcal{R}} \pi_t(r) = 1$

u = bin approximation to f : $u(r) \approx f(\xi), \;$ if $\text{bin}(\xi) = r$

ϕ_t solves Poisson eqn $\quad \left(I - Q_t + \frac{\pi_t \pi_t^T}{\|\pi_t\|_2^2}\right)\phi_t = u - \pi_t^T u \mathbb{1}$

$v_t(r)$ = rth entry of $\sqrt{Q_t(\phi_t^2) - (Q_t\phi_t)^2} \equiv \sqrt{\text{Var}_{Q_t(r,\cdot)}(\phi_t)}$

Talk organization:

1. Precise algorithm
2. Derivation of v_t & parameter choice
3. Numerical example
4. Future work

Recall $\omega_t(r) :=$ total weight in bin r at time t
$N_t(r) := \#$ of children in bin r at time t

Algorithm sketch

Choose initial population, then iterate for $t \geq 0$:

Selection. In each bin r, resample from parents w.p.p.t. their weights to get $N_t(r)$ children, each w/ weight $\frac{\omega_t(r)}{N_t(r)}$

Mutation. Evolve children via K to get next generation, update observable, set $t \leftarrow t+1$, and return to Selection

Remarks:

Total weight $= 1$, total $\#$ of particles $= N$

If initial particles are sampled from ρ_0, at time t the population law is an unbiased estimate of $\rho_0 K^t$

Details of selection very important for long time sampling

Systematic resampling: For $q_i \geq 0$,

$$\{n_i : i \in I\} = \text{resample}\left(\{q_i : i \in I\}, n\right)$$

means we draw $U \sim \text{Unif}(0,1)$, set

$$U_j = U + \frac{j-1}{n} \mod 1,$$

and then, if say $I = \{1, \ldots, m\}$,

$$n_i = \#\left\{j \in \{1, \ldots, n\} : U_j \in \left[\frac{\sum_{k=1}^{i-1} q_i}{\sum_{k=1}^{m} q_i}, \frac{\sum_{k=1}^{i} q_i}{\sum_{k=1}^{m} q_i}\right)\right\}$$

Selection step:

Let $\{\tilde{N}_t(r) : r \in \mathcal{R}\} :=$

$$\text{resample}\left(\{\omega_t(r)v_t(r) : r \in \mathcal{R}\},\ N - \sum_{r\in\mathcal{R}} \mathbb{1}_{\omega_t(r)>0}\right)$$

Define $N_t(r) = \mathbb{1}_{\omega_t(r)>0} + \tilde{N}_t(r)$

Idea: put 1 child in each occupied bin, then allocate the remainder using v_t. Thus no bin "dies"

Selection step:

Write $\{\xi_t^j, \omega_t^j\}^{j=1,\ldots,N}$ for the {particles, weights} at time t

$$\text{Let } \{\#\text{children of } \xi_t^j : \text{bin}(\xi_t^j) = r\} := \text{resample}\left(\{\omega_t^j : \text{bin}(\xi_t^j) = r\},\, N_t(r)\right)$$

Idea: Select children in each bin with probability proportional to their parents' weights

Mutation step:

Children evolve independently according to K

Q_{t+1} obtained from stochastic approximation scheme

Observable updated

$$\theta_{t+1} = \left(1 - \frac{1}{t+1}\right)\theta_t + \frac{1}{t+1}\sum_{j=1}^{N} \omega_t^j f(\xi_t^j)$$

Note: This corresponds to simple time average

Analysis:

Write $\{\xi_t^j, \omega_t^j\}^{j=1,\ldots,N}$ for the {particles, weights} at time t

$\mathcal{F}_t := \sigma(\text{random objects up to } t\text{-th selection step})$
$\hat{\mathcal{F}}_t := \sigma(\mathcal{F}_t, \text{random objects from } t\text{-th selection step})$

Consider time average of observable

$$\theta = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{N}\omega_t^j f(\xi_t^j)$$

and Doob martingale

$$D_0, \hat{D}_0, D_1, \hat{D}_1, D_2, \hat{D}_2, \ldots,$$

$$D_t := \mathbb{E}\left[\theta|\,\mathcal{F}_t\right], \; \hat{D}_t := \mathbb{E}\left[\theta|\,\hat{\mathcal{F}}_t\right].$$

Doob decomposition:

$$(\theta - \mathbb{E}[\theta])^2 = B_T + (D_0 - \mathbb{E}[\theta])^2 + \sum_{t=0}^{T-1} \underbrace{\mathbb{E}\left[(D_{t+1} - \hat{D}_t)^2 | \hat{\mathcal{F}}_t\right]}_{\text{variance from } t\text{-th mutation step}} + \sum_{t=0}^{T-1} \underbrace{\mathbb{E}\left[(\hat{D}_t - D_t)^2 | \mathcal{F}_t\right]}_{\text{variance from } t\text{-th selection step}}$$

where $\mathbb{E}[B_T] = 0$.

Some calculations show: if we replace K with Q_t,

$$\lim_{T\to\infty} \underbrace{\mathbb{E}\left[(D_{t+1} - \hat{D}_t)^2 \middle| \tilde{\mathcal{F}}_t\right]}_{\text{variance from } t\text{-th mutation step}} \approx \sum_{r\in\mathcal{R}} \frac{\omega_t(r)^2}{N_t(r)} v_t(r)^2$$

where v_t = selection value function and

$$\tilde{\mathcal{F}}_t = \sigma(\mathcal{F}_t, \{N_t(r) : r \in \mathcal{R}\}).$$

Note: The approximation $K \approx Q_t$ is uncontrolled!

Our strategy minimizes

$$\underbrace{\sum_{r\in\mathcal{R}} \frac{\omega_t(r)^2}{N_t(r)} v_t(r)^2}_{\text{mutation variance approximation}}, \quad \text{subject to } \sum_{r\in\mathcal{R}} N_t(r) = N.$$

The result is

$$N_t(r) = \frac{N\omega_t(r)v_t(r)}{\sum_{r\in\mathcal{R}} \omega_t(r)v_t(r)}.$$

Special case:

If each bin contains exactly one point of space, then:

-mutation variance is minimized (w/o approximation)
-the variance from selection is exactly zero

Here, algorithm may be "globally" optimal as $N, T \to \infty$

Synthetic variance: estimate value of variance as

$$\sigma_{\text{syn}}^2 = \sum_{r \in \mathcal{R}} \frac{\omega_0(r)^2}{N_0(r)} v_0(r)^2$$

where $\omega_0 :=$ stationary vector of Q_0.

This assumes initial conditions approximate steady state (but analogous formulas exist for other initial conditions)

Idea: Use synthetic variance to choose parameters

Numerical example: Mean first passage time

Let $(Y_t)_{t\ge 0}$ be a time discrete Markov process and

$$\tau_{loc} = \Delta t \wedge \min\{t > 0 : \mathrm{bin}(Y_t) \neq \mathrm{bin}(Y_0)\}$$

a resampling time. Given "source" ρ and "sink" R, set

$$K(\xi, d\xi') = \begin{cases} \mathbb{P}[Y_{\tau_{loc}} \in d\xi' | Y_0 = \xi], & \xi \notin R \\ \mathbb{P}[Y_{\tau_{loc}} \in d\xi' | Y_0 \sim \rho], & \xi \in R \end{cases}.$$

K is the Markov kernel that defines particle evolution.

If $\tau_R = \inf\{t \geq 0 : Y_t \in R\}$, then under mild assumptions

$$\mathbb{E}[\tau_R | Y_0 \sim \rho] = \frac{\mathbb{E}[\tau_{loc} | Y_0 \sim \mu]}{\mu(R)}.$$

where $\mu =$ stationary distribution of K.

Idea: Use algorithm to estimate the RHS, with allocation set to minimize variance of <u>denominator</u>

Example:

$(Y_t)_{t\ge 0}$ = Euler-Maruyama discretization of Brownian dynamics

time step $= 0.0001$

potential energy $V(\xi) = \begin{cases} 1 + 25(\xi - 1/2)^2, & \xi < 1/2 \\ V(\xi) = \cos(12\pi\xi), & \text{else} \end{cases}$.

$\mathcal{R}$ consists of 100 equally spaced bins in $[0,1]$

$R = [91/100, 92/100)$ is one of the bins, $\rho = \delta_{1/2}$

$N = 500$ particles and $\Delta t = 0.0002$

Initial population obtained by runs 5% as long as main simulation

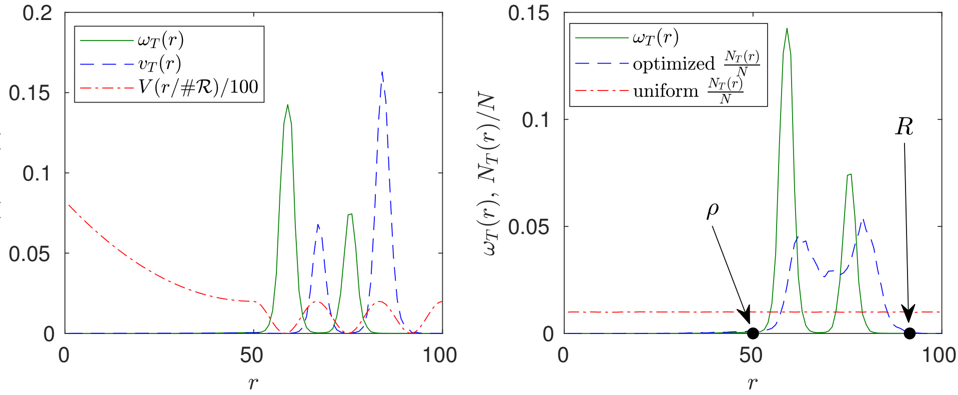

Figure: Left: bin weights $\omega_T(r)$, selection value function $v_T(r)$, and rescaled potential V. Right: bin weights vs. optimized and uniform allocation. Here ρ = source, R = sink.

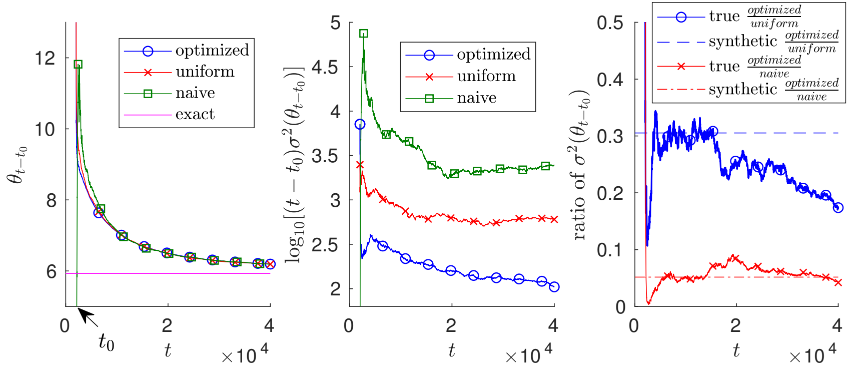

Figure: Left: convergence of observable average to MFPT. Center: scaled variances for optimized allocation, uniform allocation, and naive sampling. Right: empirical variance ratios compared to their synthetic approximations obtained before beginning simulations.

Example:

$(Y_t)_{t\geq 0} = (q_t, p_t)_{t\geq 0} =$ BBK discretization of Langevin dynamics

Time step is $\delta t = 0.001$

Muller-Brown potential energy $V : \mathbb{R}^2 \to \mathbb{R}$

Position space divided into $25^2 = 625$ equally sized rectangular bins; momenta unbinned

$R = F \times \mathbb{R}^2$, $\rho(dq, dp) = \delta_{q_0} \times \eta$, $\eta =$ Boltzmann distribution on momenta

$N = 3125$ particles and $\Delta t = 1$

Initial population obtained by runs 5% as long as main simulation

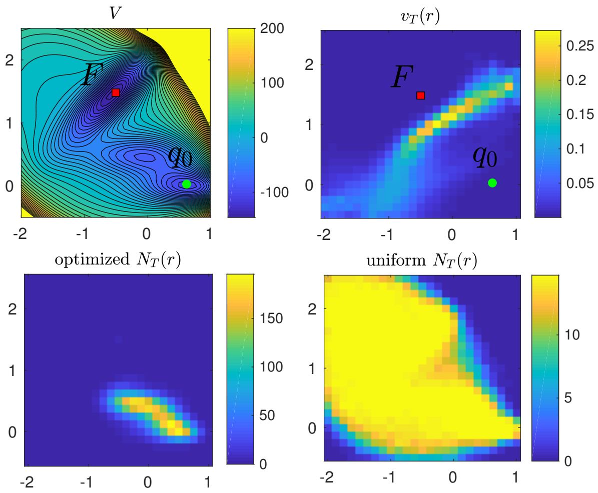

Figure: Top left: Muller-Brown potential. Top right: selection value function $v_T(r)$. Bottom: optimized and uniform allocation.

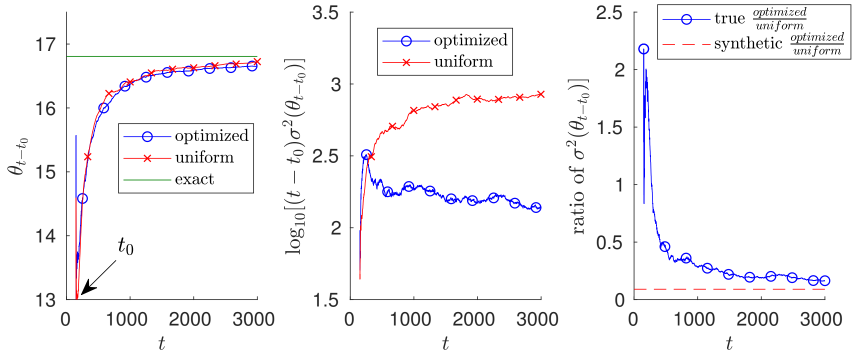

Figure: Left: convergence of observable average to MFPT. Center: scaled variances for optimized allocation, uniform allocation, and naive sampling. Right: empirical variance ratios compared to their synthetic approximations obtained *after* simulations. Here the synthetic variance was computed using the final population.

Gain over uniform allocation?

Consider $\mathcal{S}$ = set of bins r where

$$\omega_t(r)v_t(r)/\sum_{r\in\mathcal{R}}\omega_t(r)v_t(r) \gg 0$$

Let N_{unif} = avg # of particles in $\mathcal{S}$ with uniform allocation

Then the variance can be reduced by a factor of $\approx N/N_{unif}$ by using optimized instead of uniform allocation.

Note: Better estimate via synthetic variance

Future work:

Replace bin weights ω_t with stationary vector of Q_t?

Understand better local vs. global variance minimization?

Implementation on realistic systems/parameter choice?

Thanks to the organizers for the invitation!

Funding Acknowledgement:

D. Aristoff supported by NSF DMS 1522398/NSF DMS 1818726; D.M. Zuckerman supported by NIH GM115805.