Particle Methods for Rare Event Monte Carlo

Paul Dupuis

Division of Applied Mathematics
Brown University

(Thomas Dean (Oxford))

ASEAS, Arlington, VA

March, 2009

Particle methods for rare event Monte Carlo

The use of branching processes to estimate small probabilities

The use of branching processes to estimate small probabilities
Summary:

Particle methods for rare event Monte Carlo

The use of branching processes to estimate small probabilities
Summary:

- The design of such schemes was (until recently) poorly understood.

Particle methods for rare event Monte Carlo

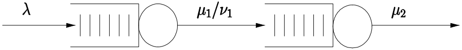
The use of branching processes to estimate small probabilities
Summary:

- The design of such schemes was (until recently) poorly understood.
- Design should be based on subsolutions to an associated HJB equation.

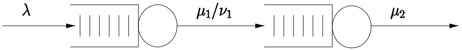
The use of branching processes to estimate small probabilities
Summary:

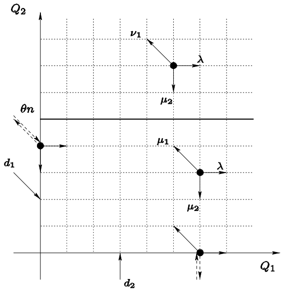
- The design of such schemes was (until recently) poorly understood.
- Design should be based on subsolutions to an associated HJB equation.
- Obtain necessary and sufficient conditions for asymptotically optimal performance.

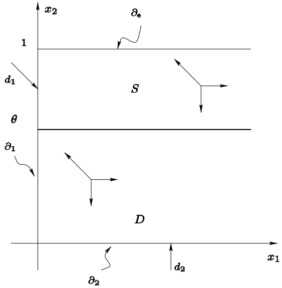
Example: A tandem queue with server slow-down (Ethernet control)



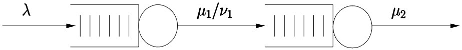
Example: A tandem queue with server slow-down (Ethernet control)

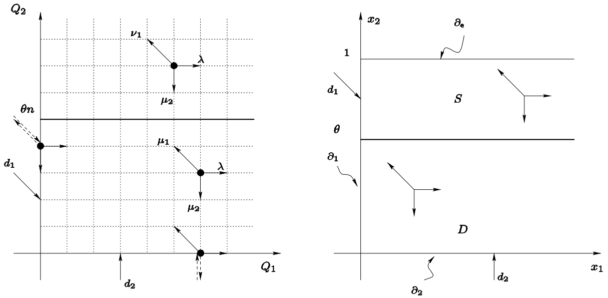






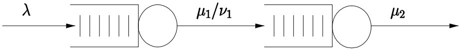
Example: A tandem queue with server slow-down (Ethernet control)

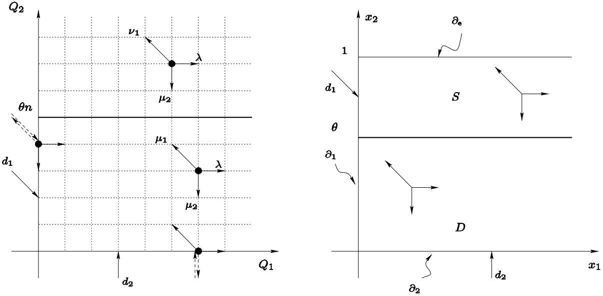




$$p_n = P\left\{Q_2 \text{ exceeds } n \text{ before } Q = (0,0) | Q(0) = (1,0)\right\}.$$

Example: A tandem queue with server slow-down (Ethernet control)





$$p_n = P\left\{Q_2 \text{ exceeds } n \text{ before } Q = (0,0) | Q(0) = (1,0)\right\}.$$

Also, analogous non-Markovian model.

Model problem and large deviation scaling

As a general Markov model one can consider iid random vector fields $\left\{v_i(y), y \in \mathbb{R}^d\right\}$, and the process

$$X_{i+1}^n = X_i^n + \frac{1}{n} v_i(X_i^n), \quad X_0^n = x.$$

Model problem and large deviation scaling

As a general Markov model one can consider iid random vector fields $\left\{v_i(y), y \in \mathbb{R}^d\right\}$, and the process

$$X_{i+1}^n = X_i^n + \frac{1}{n} v_i(X_i^n), \quad X_0^n = x.$$

Define

$$H(y,\alpha) = \log E \exp \langle \alpha, v_i(y) \rangle , \quad L(y,\beta) = \sup_{\alpha \in \mathbb{R}^d} \left[\langle \alpha, \beta \rangle - H(y,\alpha) \right]$$

$$X^n(i/n) = X_i^n, \quad \text{piecewise linear interpolation for } t \neq i/n.$$

Model problem and large deviation scaling (cont'd)

Under conditions $\{X^n(\cdot)\}$ satisfies a Large Deviation Principle with rate function

$$I_T(\phi) = \int_0^T L(\phi, \dot{\phi})dt$$

if ϕ is AC and $\phi(0) = x$, and $I_T(\phi) = \infty$ else.

Model problem and large deviation scaling (cont'd)

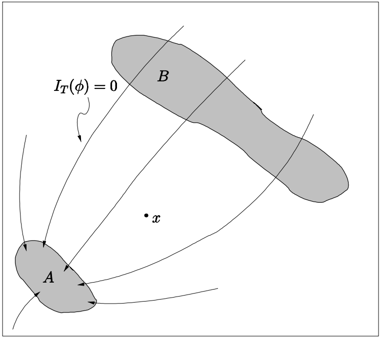
Under conditions $\{X^n(\cdot)\}$ satisfies a Large Deviation Principle with rate function

$$I_T(\phi) = \int_0^T L(\phi, \dot{\phi}) dt$$

if ϕ is AC and $\phi(0) = x$, and $I_T(\phi) = \infty$ else. Heuristically, for $T < \infty$, given ϕ, small $\delta > 0$ and large n

$$P\left\{ \sup_{0 \le t \le T} \|X^n(t) - \phi(t)\| \le \delta \right\} \approx e^{-n I_T(\phi)}.$$

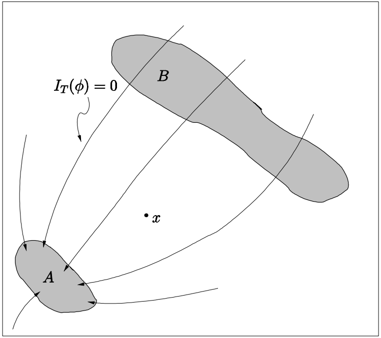
Model problem and large deviation scaling (cont'd)



Let $C = \{$ trajectories that hit B prior to A $\}$. To estimate:

$$p_n(x) = P\left\{X^n \in C | X^n(0) = x\right\}.$$

Model problem and large deviation scaling (cont'd)



Under mild conditions:

$$-\frac{1}{n}\log p_n(x) \to \inf\left\{I_T(\phi) : \phi \text{ enters } B \text{ prior to } A \text{ before } T, T < \infty\right\} = \gamma(x).$$

Some estimation generalities

1. For standard Monte Carlo we average iid copies of $1_{\{X^n \in C\}}$. One needs $K \approx e^{n\gamma}$ samples for bounded relative error [std dev$/p_n(x)$].

Some estimation generalities

1. For standard Monte Carlo we average iid copies of $1_{\{X^n \in C\}}$. One needs $K \approx e^{n\gamma}$ samples for bounded relative error [std dev$/p_n(x)$].
2. Alternative approach: construct iid random variables $\theta_1^n, \ldots, \theta_K^n$ with $E\theta_1^n = p_n(x)$ and use the unbiased estimator

$$\hat{q}_{n,K}(x) \doteq \frac{\theta_1^n + \cdots + \theta_K^n}{K}.$$

Some estimation generalities

1. For standard Monte Carlo we average iid copies of $1_{\{X^n \in C\}}$. One needs $K \approx e^{n\gamma}$ samples for bounded relative error [std dev$/p_n(x)$].
2. Alternative approach: construct iid random variables $\theta_1^n, \ldots, \theta_K^n$ with $E\theta_1^n = p_n(x)$ and use the unbiased estimator
$$\hat{q}_{n,K}(x) \doteq \frac{\theta_1^n + \cdots + \theta_K^n}{K}.$$
3. Performance determined by variance of θ_1^n, and since unbiased by $E\left(\theta_1^n\right)^2$.

Some estimation generalities

1. For standard Monte Carlo we average iid copies of $1_{\{X^n \in C\}}$. One needs $K \approx e^{n\gamma}$ samples for bounded relative error [std dev$/p_n(x)$].
2. Alternative approach: construct iid random variables $\theta_1^n, \ldots, \theta_K^n$ with $E\theta_1^n = p_n(x)$ and use the unbiased estimator

$$\hat{q}_{n,K}(x) \doteq \frac{\theta_1^n + \cdots + \theta_K^n}{K}.$$

3. Performance determined by variance of θ_1^n, and since unbiased by $E\left(\theta_1^n\right)^2$.
4. By Jensen's inequality

$$-\frac{1}{n}\log E\left(\theta_1^n\right)^2 \leq -\frac{2}{n}\log E\theta_1^n = -\frac{2}{n}\log p_n(x) \to 2\gamma(x).$$

Some estimation generalities

1. For standard Monte Carlo we average iid copies of $1_{\{X^n \in C\}}$. One needs $K \approx e^{n\gamma}$ samples for bounded relative error [std dev$/p_n(x)$].
2. Alternative approach: construct iid random variables $\theta_1^n, \ldots, \theta_K^n$ with $E\theta_1^n = p_n(x)$ and use the unbiased estimator

$$\hat{q}_{n,K}(x) \doteq \frac{\theta_1^n + \cdots + \theta_K^n}{K}.$$

3. Performance determined by variance of θ_1^n, and since unbiased by $E\left(\theta_1^n\right)^2$.
4. By Jensen's inequality

$$-\frac{1}{n}\log E\left(\theta_1^n\right)^2 \leq -\frac{2}{n}\log E\theta_1^n = -\frac{2}{n}\log p_n(x) \to 2\gamma(x).$$

5. An estimator is called *asymptotically efficient* if

$$\liminf_{n\to\infty} -\frac{1}{n}\log E\left(\theta_1^n\right)^2 \geq 2\gamma(x).$$

Splitting type schemes

- Pure branching methods (also called multi-level splitting)

Splitting type schemes

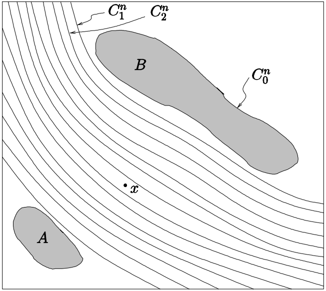
- Pure branching methods (also called multi-level splitting)
- Branching with killing [RESTART, DPR]

Splitting type schemes

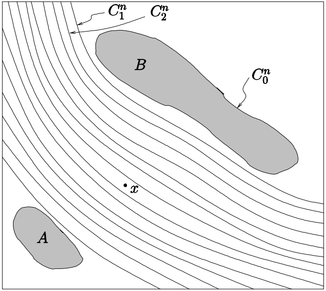
- Pure branching methods (also called multi-level splitting)
- Branching with killing [RESTART, DPR]
- Interacting particle systems (Del Moral et. al.)

Construction of splitting estimators

Pure branching. A certain number [proportional to *n*] of *splitting thresholds* C_r^n are defined which enhance migration, e.g.,

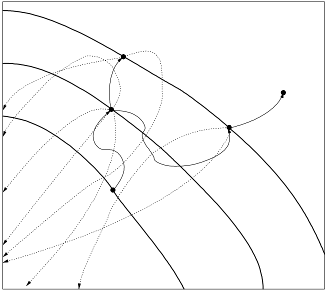


Pure branching. A certain number [proportional to n] of *splitting thresholds* C_r^n are defined which enhance migration, e.g.,

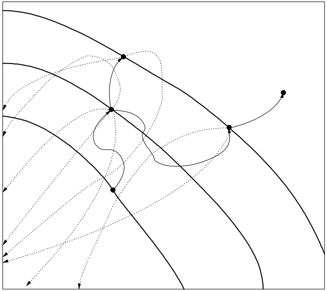


A single particle is started at x that follows the same law as X^n, but branches into a number of independent copies each time a new level is reached.

Construction of splitting estimators (cont'd)



Construction of splitting estimators (cont'd)



The number of new particles M can be random (though independent of past data), and a multiplicative weight w_i is assigned to the ith descendent, where

$$E \sum_{i=1}^{M} w_i = 1.$$

Construction of splitting estimators (cont'd)

Evolution continues until every particle has reached either A or B. Let

$$\begin{aligned} M_x^n &= \text{total number of particles generated} \\ X_j^n(t) &= \text{trajectory of } j\text{th particle,} \\ W_j^n &= \text{product of weights assigned to } j \text{ along path} \end{aligned}$$

Construction of splitting estimators (cont'd)

Evolution continues until every particle has reached either A or B. Let

$$\begin{aligned} M_x^n &= \text{total number of particles generated} \\ X_j^n(t) &= \text{trajectory of } j\text{th particle,} \\ W_j^n &= \text{product of weights assigned to } j \text{ along path} \end{aligned}$$

Then

$$\theta^n = \sum_{j=1}^{M_x^n} 1_{\{X_j^n \in C\}} W_j^n$$

Subsolutions for branching processes

Now consider the asymptotic rate of decay as a function of y:

$$\begin{aligned}\gamma(y) &= \lim_{n\to\infty} -\frac{1}{n}\log p_n(y) \\ &= \inf\left\{I_T(\phi) : \phi(0) = y, \phi \text{ enters } B \text{ prior to } A \text{ before } T, T < \infty\right\} .\end{aligned}$$

Subsolutions for branching processes

Now consider the asymptotic rate of decay as a function of y:

$$\begin{aligned}\gamma(y) &= \lim_{n\to\infty} -\frac{1}{n}\log p_n(y) \\ &= \inf\left\{ I_T(\phi) : \phi(0) = y, \phi \text{ enters } B \text{ prior to } A \text{ before } T, T < \infty \right\} .\end{aligned}$$

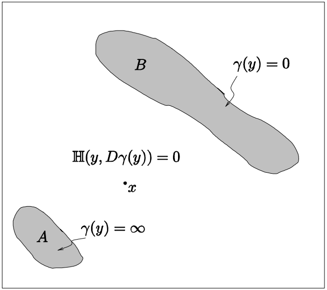
Let

$$\mathbb{H}(y,\alpha) = -H(y,-\alpha)$$

[recall $H(y,\alpha) = \log E \exp \langle \alpha, v_i(y) \rangle$].

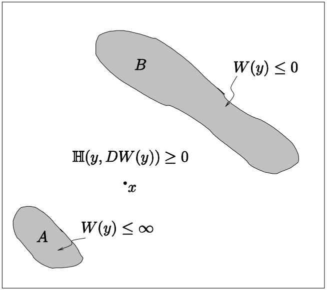
Subsolutions for branching processes (cont'd)

$\gamma(y)$ is a weak-sense solution to the PDE



Subsolutions for branching processes (cont'd)

Subsolutions should satisfy (in the viscosity sense)



Statement of results

Implementation and Performance for Pure Splitting [analogous results for RESTART, etc.]:

Implementation and Performance for Pure Splitting [analogous results for RESTART, etc.]:
Consider a continuous function W and suppose splitting levels are the level sets $\{W(y) \leq i \log EM/n\}$, where EM is the mean number of particles per split.

Statement of results

Implementation and Performance for Pure Splitting [analogous results for RESTART, etc.]:
Consider a continuous function W and suppose splitting levels are the level sets $\{W(y) \leq i \log EM/n\}$, where EM is the mean number of particles per split.

- Then the number of particles needed to construct a single sample θ_1^n grows subexponentially if and only if W is a viscosity subsolution.

Statement of results

Implementation and Performance for Pure Splitting [analogous results for RESTART, etc.]:
Consider a continuous function W and suppose splitting levels are the level sets $\{W(y) \leq i \log EM/n\}$, where EM is the mean number of particles per split.

- Then the number of particles needed to construct a single sample θ_1^n grows subexponentially if and only if W is a viscosity subsolution.

Given $u = EM$, consider particular scheme that randomizes between $\lfloor u \rfloor$ and $\lfloor u \rfloor + 1$ and uses weights $w_i = 1/u$.

Statement of results

Implementation and Performance for Pure Splitting [analogous results for RESTART, etc.]:
Consider a continuous function W and suppose splitting levels are the level sets $\{W(y) \leq i \log EM/n\}$, where EM is the mean number of particles per split.

- Then the number of particles needed to construct a single sample θ_1^n grows subexponentially if and only if W is a viscosity subsolution.

Given $u = EM$, consider particular scheme that randomizes between $\lfloor u \rfloor$ and $\lfloor u \rfloor + 1$ and uses weights $w_i = 1/u$. Then

- $$\liminf_{n \to \infty} -\frac{1}{n} \log E \left(\theta_1^n\right)^2 \geq W(x) + \gamma(x).$$

Remarks

- Subsolutions for interesting models (networks with feedback, non-Markovian systems, serve-the-longer discipline, server-slowdown dynamics, open/closed networks, path-dependent events) known.
- When available the Freidlin-Ventsel *quasipotential* can be used to construct subsolutions with optimal value.
- Subsolutions for *importance sampling* must be at least piecewise classical sense.

References

- Splitting for rare event simulation: A large deviations approach to design and analysis (T. Dean and D.), Stochastic Processes and their Applications, **119**, (2009), 562–587.
- A generalized DPR algorithm for rare event simulation (T. Dean and D.), submitted to Annals of OR.