A Generalised DPR Algorithm for Rare Event Simulation

Thomas Dean

Nomura Centre for Mathematical Finance
University of Oxford

September 2008

Acknowledgements

Joint work with Paul Dupuis, Brown University.

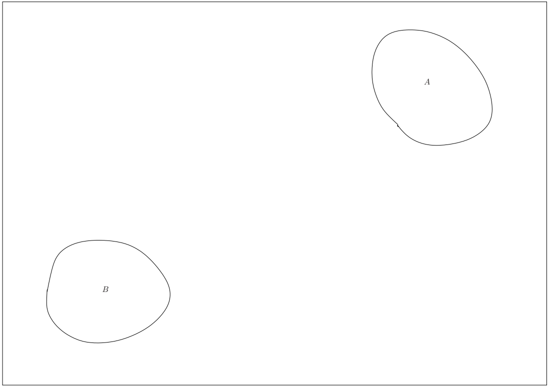
Research supported in part by National Science Foundation grants NSF-DMS-0404806 and NSF-DMS-0706003.

Overview of Talk

- Review of the RESTART and DPR algorithms.
- Description of the GDPR algorithm.
- Review of the Subsolutions Approach to Importance Sampling.
- Subsolutions Approach to Designing Splitting Algorithms.
- Examples

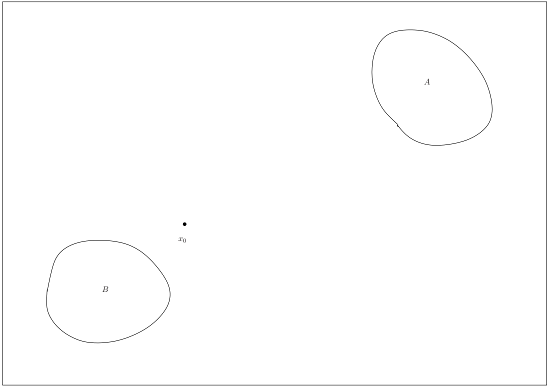
RESTART Algorithm

Consider a probability of the form $P(\tau_A < \tau_B)$.



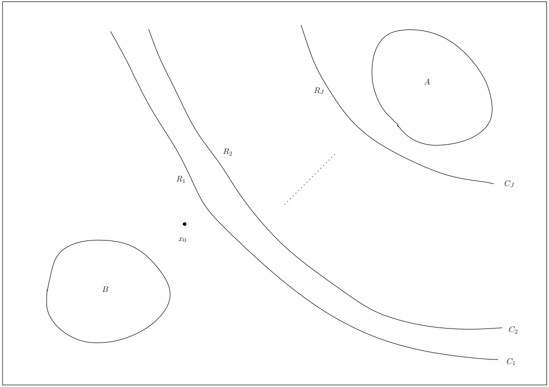
RESTART Algorithm

Consider a probability of the form $P(\tau_A < \tau_B)$.



RESTART Algorithm

Consider a probability of the form $P(\tau_A < \tau_B)$.



RESTART Algorithm

Consider a probability of the form $P(\tau_A < \tau_B)$.

RESTART Algorithm

Consider a probability of the form $P(\tau_A < \tau_B)$.

Estimate probability by $s = N / \prod_{j=1}^{J} R_j$.

Importance Functions

It is notationally cumbersome to describe algorithms using sequences of splitting thresholds and splitting rates. Instead use importance functions $V(x)$.

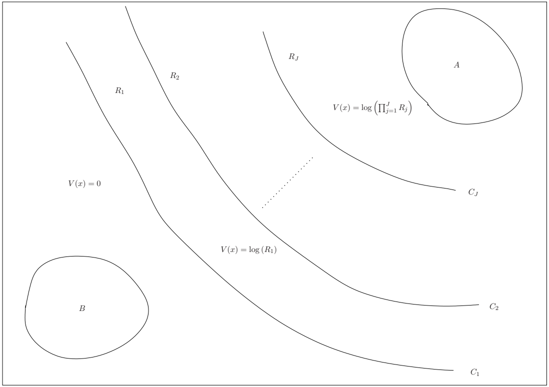
Importance Functions

It is notationally cumbersome to describe algorithms using sequences of splitting thresholds and splitting rates. Instead use importance functions $V(x)$.

- Non-negative step function.
- Sequence of nested sets $\tilde{C}_1 \supset \cdots \supset \tilde{C}_J$ such that $V(x)$ is constant on each $\tilde{C}_j/\tilde{C}_{j+1}$.
- $V(x) > V(y)$ for all x, y such that there exists j such that $x \in \tilde{C}_j$, $y \in \tilde{C}_j^c$.
- $V(x) = 0$ if $x \in \tilde{C}_1^c$.
- Will use V_j to denote the value taken by $V(x)$ on the set $\tilde{C}_j/\tilde{C}_{j+1}$.

Importance Functions

It is notationally cumbersome to describe algorithms using sequences of splitting thresholds and splitting rates. Instead use importance functions $V(x)$.



DPR Algorithm

The DPR algorithm generalises the RESTART algorithm.

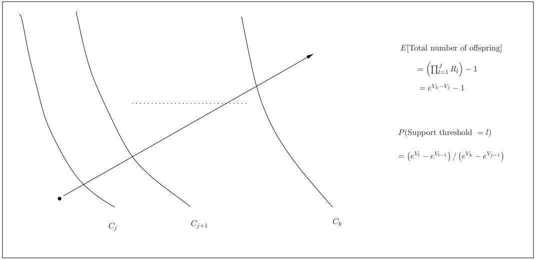
- Allows the splitting rates to take non-integer values.
- Allows the process to be able to cross more than one threshold in a single time step.

DPR Algorithm

The DPR algorithm generalises the RESTART algorithm.

- Allows the splitting rates to take non-integer values.
- Allows the process to be able to cross more than one threshold in a single time step.

Requires the number of offspring and their support thresholds to be randomised.



GDPR Algorithm

A generalised version of the DPR algorithm that can be used to estimate expected values of the form $E_x\left[\sum_{i=0}^{\tau} f(X_i)\right]$.

GDPR Algorithm

A generalised version of the DPR algorithm that can be used to estimate expected values of the form $E_x \left[\sum_{i=0}^{\tau} f(X_i)\right]$.

- Require $f(x) \geq 0$
- τ is the a.s. finite first entry time of a closed set M such that $\bar{M} = \bar{M}^{\circ}$

GDPR Algorithm

A generalised version of the DPR algorithm that can be used to estimate expected values of the form $E_x\left[\sum_{i=0}^{\tau} f(X_i)\right]$.

- Require $f(x) \geq 0$
- τ is the a.s. finite first entry time of a closed set M such that $\bar{M} = \bar{M}^\circ$

In implementation the GDPR algorithm is identical to the DPR algorithm except that the expected value is estimated by
$s = E_x\left[\sum_{i=0}^{\infty}\left(\sum_{j=1}^{N_i} f(\bar{X}_{i,j}) e^{-V(\bar{X}_{i,j})}\right)\right]$

GDPR Algorithm

A generalised version of the DPR algorithm that can be used to estimate expected values of the form $E_x\left[\sum_{i=0}^{\tau} f(X_i)\right]$.

- Require $f(x) \geq 0$
- τ is the a.s. finite first entry time of a closed set M such that $\bar{M} = \bar{M}^{\circ}$

In implementation the GDPR algorithm is identical to the DPR algorithm except that the expected value is estimated by
$s = E_x\left[\sum_{i=0}^{\infty}\left(\sum_{j=1}^{N_i} f(\bar{X}_{i,j})e^{-V(\bar{X}_{i,j})}\right)\right]$

- Can show that the GDPR algorithm terminates in a finite time a.s. and that it is unbiased.

Rare Event Simulation

Will consider the problem of estimating sequences of expected values of the form $E_n = E_x\left[\sum_{i=0}^{\tau^n} \exp(-nF(X_i^n))\right]$ where F is bounded and continuous and the sequence E_n obeys a large deviations rule $\lim -\frac{1}{n} \log E_n \to \gamma$.

Rare Event Simulation

Will consider the problem of estimating sequences of expected values of the form $E_n = E_x\left[\sum_{i=0}^{\tau^n} \exp(-nF(X_i^n))\right]$ where F is bounded and continuous and the sequence E_n obeys a large deviations rule $\lim -\frac{1}{n}\log E_n \to \gamma$.

- One has to define a sequence of importance functions V^n. Such a sequence will be called a GDPR scheme.

Rare Event Simulation

Will consider the problem of estimating sequences of expected values of the form $E_n = E_x \left[\sum_{i=0}^{\tau^n} \exp(-nF(X_i^n)) \right]$ where F is bounded and continuous and the sequence E_n obeys a large deviations rule $\lim -\frac{1}{n} \log E_n \to \gamma$.

- One has to define a sequence of importance functions V^n. Such a sequence will be called a GDPR scheme.
- Will require that $\lim \frac{1}{n} \log E_x[w^n] = 0$. Such schemes will be called stable.

Rare Event Simulation

Will consider the problem of estimating sequences of expected values of the form $E_n = E_x\left[\sum_{i=0}^{\tau^n} \exp(-nF(X_i^n))\right]$ where F is bounded and continuous and the sequence E_n obeys a large deviations rule $\lim -\frac{1}{n} \log E_n \to \gamma$.

- One has to define a sequence of importance functions V^n. Such a sequence will be called a GDPR scheme.
- Will require that $\lim \frac{1}{n} \log E_x[w^n] = 0$. Such schemes will be called stable.
- Measure performance of stable schemes with the asymptotic relative error $\lim \frac{1}{n} \log \left(E_x[(s_n)^2]/E_n^2\right)$

Rare Event Simulation

Will consider the problem of estimating sequences of expected values of the form $E_n = E_x\left[\sum_{i=0}^{\tau^n} \exp(-nF(X_i^n))\right]$ where F is bounded and continuous and the sequence E_n obeys a large deviations rule $\lim -\frac{1}{n}\log E_n \to \gamma$.

- One has to define a sequence of importance functions V^n. Such a sequence will be called a GDPR scheme.
- Will require that $\lim \frac{1}{n}\log E_x[w^n] = 0$. Such schemes will be called stable.
- Measure performance of stable schemes with the asymptotic relative error $\lim \frac{1}{n}\log\left(E_x[(s_n)^2]/E_n^2\right)$
- Goal is to find a way of constructing GDPR schemes which are stable and asymptotically optimal or close to asymptotically optimal.

Subsolutions and Importance Sampling

Under very general conditions the large deviations rate for the expected values E_n is equal to the value function of a control problem of the form

$$\inf \int_0^t L(\phi, \dot{\phi}) ds + F(\phi_t)$$

Subsolutions and Importance Sampling

Under very general conditions the large deviations rate for the expected values E_n is equal to the value function of a control problem of the form

$$\mathcal{J}(y,z) = \inf_{\phi:\phi(0)=y;\phi(T)=z;\phi(s)\notin M, s\in(0,T);T<\infty} \int_0^T L(\phi(s),\dot{\phi}(s))ds.$$

Subsolutions and Importance Sampling

Under very general conditions the large deviations rate for the expected values E_n is equal to the value function of a control problem of the form

$$\inf_y \mathcal{J}(x, y) + F(y)$$

Subsolutions and Importance Sampling

Under very general conditions the large deviations rate for the expected values E_n is equal to the value function of a control problem of the form

$$\inf_y \mathcal{J}(x, y) + F(y)$$

- A standard heuristic is to twist the sampling distribution so that its mean path becomes equal to the 'most significant' path in the large deviations analysis.

Subsolutions and Importance Sampling

Under very general conditions the large deviations rate for the expected values E_n is equal to the value function of a control problem of the form

$$\inf_y \mathcal{J}(x, y) + F(y)$$

- A standard heuristic is to twist the sampling distribution so that its mean path becomes equal to the 'most significant' path in the large deviations analysis.
- Dupuis and Wang observed that the control representation of the LDP suggests a natural 'feedback' importance sampling change of measure.

Subsolutions and Importance Sampling

Under very general conditions the large deviations rate for the expected values E_n is equal to the value function of a control problem of the form

$$\inf_y \mathcal{J}(x, y) + F(y)$$

- A standard heuristic is to twist the sampling distribution so that its mean path becomes equal to the 'most significant' path in the large deviations analysis.
- Dupuis and Wang observed that the control representation of the LDP suggests a natural 'feedback' importance sampling change of measure.
- In general one does not need to know the value function of the control problem, W, subsolutions are sufficient.

Generating Functions

A generating function $U(x)$ is a uniformly continuous function that is bounded from below.

Given a generating function U a sequence of importance functions V^n is produced in the following manner.

Generating Functions

A generating function $U(x)$ is a uniformly continuous function that is bounded from below.

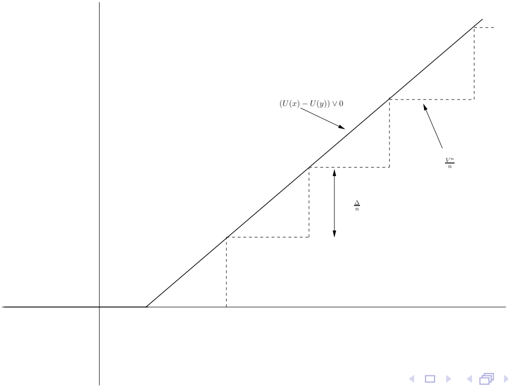
Given a generating function U a sequence of importance functions V^n is produced in the following manner.

- Choose a positive number Δ.
- For each n let V^n/n be equal to the step function whose value at any point y is equal to the largest integer multiple of Δ/n less than or equal to $(U(x) - U(y)) \vee 0$.

Generating Functions

A generating function $U(x)$ is a uniformly continuous function that is bounded from below.

Given a generating function U a sequence of importance functions V^n is produced in the following manner.



Assumptions

Continuous time interpolant $Y^n(t) = X_i^n,\ t \in [\frac{i}{n}, \frac{i+1}{n})$

Condition

For every $T \in (0, \infty)$ the sequence $\{Y^n, n = 1, 2, \ldots\}$ satisfies a large deviation principle (LDP) on $\mathcal{D}([0, T] : D)$ with a rate function of the form

$$\int_0^T L(\phi(s), \dot{\phi}(s))ds$$

if $\phi \in \mathcal{D}([0, T] : D)$ is absolutely continuous and ∞ otherwise. This LDP is uniform with respect to initial conditions in compact sets.

Assumptions

Condition

For any compact $\kappa \subset D$

$$\limsup_{T\to\infty} \limsup_{n\to\infty} \sup_{x\in\kappa} -\frac{1}{n} \log P_x \left\{ \tau^n / n \geq T \right\} = -\infty$$

and for any $l \in \mathbb{Z}^+$

$$\limsup_{n\to\infty} \sup_{x\in\kappa} \frac{1}{n} \log E_x \left[(\tau^n)^l \right] < \infty.$$

Assumptions

Condition

1. *Let* $\varepsilon > 0$*. Given any compact set* $K \subset D$*, there is* $\delta > 0$ *such that if* $x, y \in K \cap (M^\circ)^c$ *satisfy* $\|x - y\| \leq \delta$*, then there is* $\sigma \leq \varepsilon$ *and a trajectory* ϕ *connecting* x *to* y *such that* $\phi(r) \notin M$ *for all* $r \in (0, \sigma)$ *and* $\int_0^\sigma L(\phi(s), \dot{\phi}(s))ds \leq \varepsilon$*.*
2. *Let* $T < \infty$ *and a bounded and continuous function* H *be given. Consider any sequence of times* $i_n \leq \tau^n \wedge \lfloor nT \rfloor$ *such that* $i_n/n \to t \leq T$*. Then*

$$\limsup_{n\to\infty} \frac{1}{n} \log E_x e^{-nH(X^n_{i_n})} \leq -\inf \left[\int_0^t L(\phi(r), \dot{\phi}(r))dr + H(\phi(t)) \right],$$

where the infimum is over all ϕ *that satisfy* $\phi(r) \notin M$ *for* $r \in (0, t)$*.*

Subsolutions Approach to Splitting

Definition

A continuous function $G(y)$ is a subsolution if $G(y) - G(z) \leq \mathcal{J}(y, z)$ for all $y, z \in D$.

Subsolutions Approach to Splitting

Definition

A continuous function $G(y)$ is a subsolution if $G(y) - G(z) \leq \mathcal{J}(y, z)$ for all $y, z \in D$.

Theorem

Suppose that the generating function U is a subsolution. Then

$$\lim_{n\to\infty} -\frac{1}{n} \log E_x \left[(s^n)^2 \right] = \inf_{y \in D} \left[\mathcal{J}(x, y) + (U(x) - U(y)) \vee 0 + 2F(y) \right].$$

Subsolutions Approach to Splitting

Definition

A continuous function $G(y)$ is a subsolution if
$G(y) - G(z) \leq \mathcal{J}(y, z)$ for all $y, z \in D$.

Theorem

If the generating function $U(x)$ is a subsolution then

$$\lim_{n \to \infty} \frac{1}{n} \log E_x \left[w^n \right] = 0.$$

If the generating function $U(x)$ is not a subsolution then there exists some initial condition y such that

$$\liminf_{n \to \infty} \frac{1}{n} \log E_y \left[w^n \right] > 0.$$

Finding Subsolutions

- For many problems of interest viscosity subsolutions to the HJB PDE corresponding to the LDP control problem are also subsolutions in the above sense.

Finding Subsolutions

- For many problems of interest viscosity subsolutions to the HJB PDE corresponding to the LDP control problem are also subsolutions in the above sense.
- For a viscosity subsolution $U(y) \leq F(y)$ for all y. Follows that the asymptotic relative error for the related scheme is less than or equal to $W(x) - U(x)$.

Finding Subsolutions

- For many problems of interest viscosity subsolutions to the HJB PDE corresponding to the LDP control problem are also subsolutions in the above sense.
- For a viscosity subsolution $U(y) \leq F(y)$ for all y. Follows that the asymptotic relative error for the related scheme is less than or equal to $W(x) - U(x)$.
- Can often construct asymptotically or nearly asymptotically optimal schemes for more complicated problems using collections of simple subsolutions.

Example

- Consider a tandem Jackson network with arrival and service rates $\mu_1, \mu_2 > \lambda$.

Example

- Consider a tandem Jackson network with arrival and service rates $\mu_1, \mu_2 > \lambda$.
- Let $A = \{y : y_1 + y_2 \geq 1\}$

Example

- Consider a tandem Jackson network with arrival and service rates $\mu_1, \mu_2 > \lambda$.
- Let $A = \{y : y_1 + y_2 \geq 1\}$
- Estimate probability that a trajectory of rescaled process starting from $(\frac{1}{n}, 0)$ reaches A before $(0, 0)$ and stationary measure of set A.

Example

- Consider a tandem Jackson network with arrival and service rates $\mu_1, \mu_2 > \lambda$.
- Let $A = \{y : y_1 + y_2 \geq 1\}$
- Estimate probability that a trajectory of rescaled process starting from $(\frac{1}{n}, 0)$ reaches A before $(0,0)$ and stationary measure of set A.
- HJB PDE

$$\mathbb{H}(D\bar{V}(y)) \geq 0$$

$$\mathbb{H}(D\bar{V}(y) - a(0,1)) \vee \langle (0,1), -D\bar{V}(y) + a(0,1) \rangle \geq 0, y_1 = 0$$

$$\mathbb{H}(D\bar{V}(y) - a(1,0)) \vee \langle (1,-1), -D\bar{V}(y) + a(1,0) \rangle \geq 0, y_2 = 0$$

$$\bar{V}(y) \leq 0, y \in A$$

$$\mathbb{H}(y) = -\log[\lambda e^{-y_1} + \mu_1 e^{(y_1 - y_2)} + \mu_2 e^{y_2}]$$

Example

- Consider a tandem Jackson network with arrival and service rates $\mu_1, \mu_2 > \lambda$.
- Let $A = \{y : y_1 + y_2 \geq 1\}$
- Estimate probability that a trajectory of rescaled process starting from $(\frac{1}{n}, 0)$ reaches A before $(0, 0)$ and stationary measure of set A.
- The function $U(y) = \left([\log \frac{\mu_1}{\lambda} \wedge \log \frac{\mu_2}{\lambda}] \cdot [1 - y_1 - y_2]\right) \vee 0$ is a subsolution, further $U_{(0,0)} = W_{(0,0)}$ and so the resulting schemes are asymptotically optimal.

Results

$$\mu_1 = 4.5,\ \mu_2 = 4.5,\ \lambda = 1$$

n	30	40	50
Theoretical Value	2.63×10^{-18}	1.03×10^{-24}	3.80×10^{-31}
Estimate	2.63×10^{-18}	1.06×10^{-24}	3.83×10^{-31}
Std. Err.	0.08×10^{-18}	0.04×10^{-24}	0.15×10^{-31}
95% C.I.	$[2.47, 2.79] \times 10^{-18}$	$[0.99, 1.14] \times 10^{-24}$	$[3.54, 4.13] \times 10^{-31}$
Time Taken (s)	3	6	8

Table: Hitting Probabilities

n	20	30	40
Theoretical Value	1.43×10^{-12}	6.16×10^{-19}	2.39×10^{-25}
Estimate	1.55×10^{-12}	6.39×10^{-19}	2.34×10^{-25}
Std. Err.	0.32×10^{-12}	1.54×10^{-19}	0.57×10^{-25}
95% C.I.	$[0.82, 2.28] \times 10^{-12}$	$[3.38, 9.41] \times 10^{-19}$	$[1.22, 3.46] \times 10^{-25}$
Time Taken (s)	0.4	0.6	1

Table: Stationary Measures

References 1

M. Villen-Altamirano and J. Villen-Altamirano. Analysis of RESTART simulation: Theoretical basis and sensitivity study. *European Transactions on Telecommunications*, 13:373–385, 2002.

Z. Haraszti and J. K. Townsend. The theory of direct probability redistribution and its application to rare event simulation. *ACM Transactions on Modelling and Computer Simulation*, 9:105–140, 1999.

References 2

T. Dean and P. Dupuis. The design and analysis of a generalised DPR algorithm for rare event simulation. preprint.

P. Dupuis and R. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, New York, 1997.

P. Dupuis and H. Wang. Subsolutions of an Isaacs equation and efficient schemes for importance sampling, *Math. of Op. Res.*, 32:1–35, 2007.