On the subsolution approach to efficient importance sampling

Boualem Djehiche
KTH Royal Institute of Technology, Stockholm

Based on joint work with Henrik Hult and Pierre Nyquist

We explore the connection between the following areas:

- Large Deviations
- Hamilton-Jacobi equations (some aspects of weak KAM theory)
- Rare event simulations: Importance Sampling; Multiple-Splitting, Sequential Monte Carlo; MCMC.

General problem

Let $X^n = \{X^n(t), t \geq 0\}$ be a sequence of Markov processes on $(\Omega, \mathcal{F}, P^n)$.
Given $A \in \mathcal{F}$, we want to compute

$$p_n := P^n\{X^n(1) \in A\}$$

using some unbiased estimator Z_n with minimal variance.

Using Monte Carlo simulations, one generates independent copies of X^n, under P^n, and estimates p_n by empirical means

$$p_n^M := \frac{1}{M}\sum_{i=1}^{M} \mathbb{1}_{\{X^{n,i}(1)\in A\}},$$

which are unbiased: $E[p_n^M] = p_n$. But, the variance of a single replication is

$$Var(\mathbb{1}_{\{X^{n,i}(1)\in A\}}) = p_n - p_n^2 \simeq p_n \gg p_n^2,$$

for p_n small.

A very large sample size is needed for an accurate estimation of p_n.

• Importance sampling is based on the idea that the Markov process X^n is generated from a different probability $\bar{P}^n$, and p_n is estimated by a sample mean of independent r.v. of the form

$$Z_n := \frac{dP^n}{d\bar{P}^n}(X^n)\mathbb{1}_{\{X^n(1)\in A\}}.$$

We have

$$E^{\bar{P}^n}[Z_n] = p_n \qquad \text{(Unbiasedness)}$$

and

$$E^{\bar{P}^n}[Z_n^2] = E^{P^n}[Z_n].$$

Hence,

$$Var(Z_n) = E^{\bar{P}^n}[Z_n^2] - p_n^2 = E^{P^n}[Z_n] - p_n^2.$$

It is only the second moment that depends on the sampling probability $\bar{P}^n$.

• $\bar{P}^n$ may be chosen such that $E^{P^n}[Z_n] \sim p_n^2$ which is much better than p_n if a Monte Carlo approach is used.

This is achieved if we e.g. impose the following asymptotic efficiency rule: Whenever

$$\lim_{n\to\infty} -\frac{1}{n}\log p_n = \gamma, \qquad \text{(Large Deviations Principle)}$$

Z_n should satisfy

$$\liminf_{n\to\infty} -\frac{1}{n}\log E^{\bar{P}^n}[Z_n^2] \geq 2\gamma, \qquad \text{(Asymptotic Efficiency)}$$

since, by Jensen's inequality, $E^{\bar{P}^n}[Z_n^2] \geq p_n^2$, we have

$$\limsup_{n\to\infty} -\frac{1}{n}\log E^{\bar{P}^n}[Z_n^2] \leq 2\gamma.$$

2γ is called the 'optimal decay rate'.

Noting that

$$E^{\bar{P}^n}[Z_n^2] = E^{P^n}[Z_n].$$

The above asymptotic efficiency criterion states that the 'optimal' sampling probability $\bar{P}^{*n}$ is the one which satisfies

$$\liminf_{n\to\infty} -\frac{1}{n}\log E^{P^n}\left[\frac{dP^n}{d\bar{P}^{*n}}(X^n)\mathbb{1}_{\{X^n(1)\in A\}}\right] \geq 2\gamma, \qquad (1)$$

whenever, p_n satisfies the Large Deviations limit:

$$\lim_{n\to\infty} -\frac{1}{n}\log P^n\{X^n(1)\in A\} = \gamma.$$

Importance sampling seen as a control problem

The optimal sampling probability $\bar{P}^{*n}$ can be viewed as an 'optimal control' variable and

$$W^n(t,x) := -\frac{1}{n} \inf_{\bar{P}^n} \log E_{x,t}^{P^n}\left[\frac{dP^n}{d\bar{P}^n}(X^n)\mathbb{1}_{\{X^n(1)\in A\}}\right]$$

conditional upon $X^n(t) = x, \; t < 1$ (Note that $W^n(1,x) = 0$), is the value-function of an associated control problem which satisfies a Hamilton-Jacobi type equation.

A limiting argument can be used to check that

$$W(t,x) := \lim_{n\to\infty} W^n(t,x)$$

is the value function of a limiting control problem and solves (in an appropriate sense) the Hamilton-Jacobi type equation:

$$W_t(t,x) - 2H(x, -\frac{DW}{2}(t,x)) = 0, \qquad W(1,x) = 0, \quad \text{for } x \in A,$$

where, the 'Hamiltonian' $H(x,p)$ is the Legendre transform of the local LDP rate function, L, of the Markov process X^n under P^n:

$$H(x,p) = \sup_q \{ < p, q > - L(x,q) \}.$$

To find an explicit solution is usually "out of reach" except for simple cases.

The subsolution approach

The subsolution approach to efficient importance sampling suggested by Paul Dupuis and Hui Wang (2007) is based on the following observation:

Theorem (Dupuis and Wang (2007))

$$\lim_{n\to\infty} -\frac{1}{n} \inf_{\bar{P}^n} \log E_{x,t}^{P^n} \left[\frac{dP^n}{d\bar{P}^n}(X^n) \mathbb{1}_{\{X^n(1)\in A\}} \right] \geq \bar{W}(t,x).$$

In particular, starting at $(x=0, t=0)$,

$$\lim_{n\to\infty} -\frac{1}{n} \inf_{\bar{P}^n} \log E^{P^n} \left[\frac{dP^n}{d\bar{P}^n}(X^n) \mathbb{1}_{\{X^n(1)\in A\}} \right] \geq \bar{W}(0,0),$$

where, $\bar{W}(x,t)$ satisfies

$$\begin{cases} \bar{W}_t(t,x) - 2H(x, -\frac{D\bar{W}}{2}(t,x)) \geq 0, \\ \bar{W}(1,x) \leq 0, \qquad \text{for } x \in A. \end{cases} \tag{2}$$

$\bar{W}$ is a so-called subsolution of our Hamilton-Jacobi equation:

$$W_t(t,x) - 2H(x, -\frac{DW}{2}(t,x)) = 0, \qquad W(1,x) = 0, \quad \text{for } x \in A,$$

Dupuis and Wang's algorithm

• By constructing a subsolution $\bar{W}$ to the HJ equation a change of measure (using Girsanov or Esscher transforms) can be based on $D\bar{W}(x,t)$ and the performance (measured as the exponential decay rate of the second moment) of the corresponding algorithm is given by $\bar{W}(0,0)$.

• For a pure jump process with intensity $\lambda(x)$ under P, it will have intensity

$$\tilde{\lambda}(t,x) := \lambda(x) \exp\{-\frac{D\bar{W}}{2}(t,x)\}$$

under the new measure $\bar{P}^*$.

• For a diffusion process with drift $b(t,x)$ under P, it will have drift

$$\tilde{b}(t,x) := b(t,x) - \frac{D\bar{W}}{2}(t,x)$$

under the new measure $\bar{P}^*$.

• The importance sampling algorithm based on a subsolution $\bar{W}$ is asymptotically efficient if

$$\bar{W}(0,0) = 2\gamma.$$

A systematic construct such a subsolution

We suggest a systematic way to construct such a subsolution of

$$W_t(t,x) - 2H(x, -\frac{1}{2}DW(t,x)) = 0, \qquad W(1,x) = g(x).$$

We will do it for the function $V(t,x) := \frac{1}{2}W(1-t,x)$ which solves the forward HJ equation:

$$V_t(t,x) + H(x, DV(t,x)) = 0, \qquad V(0,x) = g(x).$$

Recalling that

$$H(x,p) = \sup_q \{< p, q > - L(x,q)\},$$

V is the value-function of the variational problem

$$V(x,t) = \inf\left\{ g(\psi(0)) + \int_0^t L(\psi(s), \dot{\psi}(s) ds, \; \psi \text{ abs. cont. } \psi(t) = x \right\}$$

For $c \in \mathbb{R}$, we consider the 'viscosity' (sub)-solution of the stationary HJ equation (weak KAM theory by Fathi)

$$H(y, DS(y)) = c, \quad y \in \mathbb{R}^n. \tag{3}$$

Fathi's weak KAM theory suggests that:

• There is *one value* c_H (Mañé critical value) which depends on the Hamiltonian H for which there is a 'viscosity' solution of the stationary HJ equation

$$H(y, DS(y, c_H)) = c_H.$$

• For $c \geq c_H$, there is a viscosity' subsolution of $H(y, DS(y)) = c$.

• For $c < c_H$, there is *no* viscosity' subsolution of $H(y, DS(y)) = c$.

The Mañé critical value c_H is the smallest c for which (3) admits a viscosity subsolution.
We have

$$c_H \geq sup_x \inf_p H(x, p).$$

But, in a few cases it holds that $c_H = sup_x \inf_p H(x, p)$.

Critical diffusion process. Let $U : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a potential function and $b(x) = -DU(x)$. The infinitesimal generator of the critical diffusion is

$$\mathcal{A}f(x) = \frac{1}{2}\Delta f(x) + < b(x), Df(x) >$$

and the Hamiltonian is

$$H(x, p) := e^{-<x,p>}\mathcal{A}e^{<x,p>} =< b(x), p > + \frac{1}{2}|p|^2.$$

Then, since $H(x, Du) = H(x, -b) = -\frac{1}{2}|b(x)|^2, \quad x \in \mathbb{R}^n$,

$$c_H = sup_x \inf_p H(x, p) = -\frac{1}{2} \inf_x |b(x)|^2.$$

In particular, if $DU(x) = 0$ for some x, then $c_H = 0$.

Birth-and-Death process. The infinitesimal generator of the Birth-and-Death process defined over an interval $(a, b) \subset \mathbb{R}$ is

$$\mathcal{A}f(x) = \lambda(x)(f(x+1) - f(x)) + \mu(x)(f(x-1) - f(x))$$

and the Hamiltonian is

$$H(x, p) := e^{-xp}\mathcal{A}e^{xp} = \lambda(x)(e^{p} - 1) + \mu(x)(e^{-p} - 1).$$

In this case

$$c_H = sup_x \inf_p H(x, p) = -\inf_x(\sqrt{\mu(x)} - \sqrt{\lambda(x)})^2,$$

provided that the strictly positive intensities λ and μ satisfy

$$\int_a^b \log\left(\mu(x)/\lambda(x)\right) dx < \infty.$$

Because, the function $U(x) := \int_a^x \log\left(\mu(z)/\lambda(z)\right) dz$ satisfies

$$H(x, DU(x)) = -(\sqrt{\mu(x)} - \sqrt{\lambda(x)})^2 \leq -\inf_x(\sqrt{\mu(x)} - \sqrt{\lambda(x)})^2.$$

Pure birth process.

Let $\lambda : [0,\infty)^n \longrightarrow [0,\infty)^n$. The Hamiltonian of the pure birth process defined over $\mathbb{R}^n$ is

$$H(x,p) = \sum_{j=1}^{n} \lambda_j(x)(e_j^p - 1).$$

In this case

$$c_H = sup_x \inf_p H(x,p) = -\inf_x \sum_{j=1}^{n} \lambda_j(x) := -\lambda^*.$$

This is due to the fact that for any $c > -\lambda^*$ and $\alpha \leq \log(1 + c/\lambda^*)$, the function $\alpha \sum_{i=1}^{n} x_i$ is a subsolution to $H(x, DS(x)) = c$.

Properties of the Mañé potential

To each $c \in \mathbb{R}$ we associate the Mañé potential at x associated with L:

$$S^c(x,y) = \inf \left\{ \int_0^t \left(c + L(\psi(s), \dot{\psi}(s)) \right) ds, \ \ \psi \text{ abs. cont. } \psi(0) = x, \ \psi(t) = y, t > 0 \right\}$$

$S^c(x,y)$ enjoys the following properties:

- For each $x \in \mathbb{R}^n$, $S^c(x,x) = 0$,
- $c \longrightarrow S^c(x,y)$ is nondecreasing and satisfies the triangle inequality:

$$S^c(x,y) \leq S^c(y,z) + S^c(y,z), \quad x,y,z \in \mathbb{R}^n.$$

Let $\mathcal{S}_x^c$ be the collection of all viscosity subsolutions of $H(y, DS(y)) = c$ that vanish at x.

The following result summarizes the relationship between the Mañé critical value and Mañé potential.

Proposition Suppose $c \geq c_H$ and $x \in \mathbb{R}^n$.

• Suppose $S^c > -\infty$. Then for each $x \in \mathbb{R}^n$ the function $y \longmapsto S^c(x, y)$ is a viscosity subsolution to $H(y, DS(y)) = c$ on $\mathbb{R}^n$ and a viscosity solution on $\mathbb{R}^n \backslash \{x\}$.

• For each $y \in \mathbb{R}^n$

$$S^c(x, y) = \sup_{S \in \mathcal{S}_x^c} S(y),$$

• For each $x \in \mathbb{R}^n$,

$$U(t, y; x) := \sup_{c \geq c_H} \{S^c(x, y) - ct\}$$

is a viscosity solution of the HJB equation

$$V_t(t, y) + H(y, DV(t, y)) = 0, \quad \text{on } (0, \infty) \times \mathbb{R}^n \backslash \{x\}.$$

A duality result

Duality (in time) Theorem For every $x, y \in \mathbb{R}^n$,

$$S^c(x,y) = \inf_{t>0}\{U(t,y;x) + ct\}, \quad c \geq c_H$$

and

$$U(t,y;x) = sup_{c \geq c_H}\{S^c(x,y) - ct\}.$$

A min-max representation of the solution of the HJB equation

Theorem. For all $(t, y) \in [0, \infty) \times \mathbb{R}^n$

$$V(t, y) = \inf_x \sup_{c \geq c_H} \{g(x) + S^c(x, y) - ct\} . \qquad (4)$$

Remark In general it is not possible to interchange the inf and sup in the min-max representation.

The Hopf-Lax-Oleinik formula as special case

The classical 'Hopf-Lax-Oleinik' formula

$$V(t,y) = \inf_x \left\{ g(x) + tL\left(\frac{y-x}{t}\right)\right\}.$$

when the Hamiltonian H is 'state independent' i.e. $H(x,q) := H(q)$ is a special case our result. Indeed, this follows from the fact that

Lemma. For all $x, y \in \mathbb{R}^n$, it holds that

$$\sup_{c \geq c_H} \{S^c(x,y) - ct\} = tL\left(\frac{y-x}{t}\right).$$

Hence,

$$V(t,y) = \inf_x \left\{ g(x) + tL\left(\frac{y-x}{t}\right)\right\} = \inf_x \sup_{c \geq c_H} \{g(x) + S^c(x,y) - ct\}.$$

An algorithm to construct a subsolution

(i) For arbitrary c, find a conservative vector field $\alpha(\cdot, c)$ that solves

$$H(x, \alpha(x; c)) = c.$$

(ii) Find a potential function $u(\cdot; c)$, i.e.,

$$Du(\cdot; c) = \alpha(\cdot; c).$$

(iii) That the function

$$W(t, x; c) = 2 \inf_y \{g(y) + u(y; c) - u(x; c) - c(1 - t)\}$$

satisfies $\bar{W}_t(t, x) - 2H(x, -\frac{1}{2} D\bar{W}(t, x)) \geq 0, \quad \bar{W}(1, x) \leq 2g(x).$

i.e. a subsolution of
$W_t(t, x) - 2H(x, -\frac{1}{2} D\bar{W}(t, x)) = 0, \quad W(1, x) = 2g(x).$

(iv) The parameter $c^* := \arg\max \bar{W}(0, 0; c)$ satisfies

$$\bar{W}(0, 0; c^*) = 2\gamma. \qquad \text{(Asymptotic efficiency)}$$

Application to a simple credit risk model with contagion

- A credit portfolio consisting of n obligors is divided into d groups.

- Let $w_1, \ldots, w_d$ be the fraction of obligors in each group, $w_j > 0, \sum_{i=1}^d w_i = 1$.

- Let $Q^n(t) = (Q_1^n(t), \ldots, Q_d^n(t))$ be the number of defaults in each group, at time $t \in [0, T]$. Q^n is modeled as a continuous time pure birth process with intensity $n\lambda(Q^n(t)/n)$ where

$$\lambda(x) = (\lambda_1(x), \ldots, \lambda_d(x)), \quad \lambda_j(x) = a(w_j - x_j)e^{b\sum_{k=1}^d x_k},$$

where, $b > 0$ measures the 'strength' of the contagion from the rest of the population.

let $X^n(t) := Q^n(t)/n$.

- The objective is to compute the probability that at least a fraction z has defaulted by time $t = 1$:

$$p_n := P\left(\sum_{j=1}^d Q^n(1) \geq nz\right) = P\left(\sum_{j=1}^d X^n(1) \geq z\right).$$

• This model has been studied by R. Carmona and S. Crépey (Int. J. Theor. Appl. Finance, 2010).

• They apply a state-independent change of measure and show by numerical experiments that importance sampling performs poorly in the presence of contagion ($b > 0$).

• We will solve the problem by means of the subsolution approach. We will show that an appropriate change of measure which leads to an asymptotically efficient algorithm is in fact state-dependent and strongly related to the way the subsolution is constructed, in particular c^*.

The process X^n satisfies an LDP (see the book by Dembo and Zeitouni)

$$\lim_{n\to\infty} -\frac{1}{n}\log P_x\left(X^n \in A\right) \sim \inf\left\{\int_0^T L(\psi(t),\dot{\psi}(t))dt,\ \psi \text{ abs. cont.}\, \psi(0)=x, \psi \in A\right\},$$

where,

$$L(x,\beta) = <\beta, \log\frac{\beta}{\lambda(x)}> - <\beta-\lambda(x), \mathbb{1}>.$$

In particular, for $A := \{x,\ \sum_{i=1}^d x_i < z\}$,

$$\begin{aligned}
\lim_{n\to\infty} -\tfrac{1}{n}\log p_n &= \lim_{n\to\infty} -\tfrac{1}{n}\log P\left(\textstyle\sum_{j=1}^d X^n(1) \geq z\right) \\
&= \inf\left\{\textstyle\int_0^1 L(\psi(t),\dot{\psi}(t))dt,\ \psi \quad \text{absol. cont. } \psi(0)=0, \psi(1) \notin A\right\} \\
&:= \gamma.
\end{aligned}$$

- For this problem the associated Hamilton-Jacobi equation reads:

$$W_t(t,x) - 2H(x, -DW(t,x)/2) = 0, \quad W(1,x) = 0 \quad \text{for} \quad \sum_{i=1}^{d} x_i \geq z.$$

where, the Hamiltonian is

$$H(x,q) := \sup_{\beta}\{< \beta, q > -L(x,\beta)\} = \sum_{i=1}^{d} \lambda_j(x)(e^{q_j} - 1).$$

Clearly, $\bar{W}(x,t;c^*)$ is the appropriate subsolution where,

$$c^* := \arg\max \bar{W}(0,0;c)$$

satisfies

$$\bar{W}(0,0;c^*) = 2\gamma.$$

- The corresponding importance sampling algorithm uses the intensity

$$\tilde{\lambda}(t,x) := \lambda(x) \exp\{-D\bar{W}(t,x;c^*)/2\}.$$

The one dimensional case: Finding a subsolution

- Given c, find $u(x,c)$ which solves $Du(x,c) = \alpha$, where

$$H(x,\alpha) := \lambda(x)(e^{\alpha} - 1) = c.$$

A solution is $\quad u(x,c) = \int_x^z \log\left(1 + \frac{c}{\lambda(y)}\right) dy.$

- The function

$$\bar{W}(t,x;c) := u(x,c) - c(1-t)$$

is a subsolution of our HJ equation:

$$\bar{W}_t(t,x) - 2H(x, -D\bar{W}(t,x)/2) \le 0, \quad \bar{W}(1,x) \ge 0, \quad \text{for} \quad x \ge z.$$

- (to be solved numerically, for concrete cases)

$$c^* = \arg\max \bar{W}(0,0;c) = \arg\max \int_0^z \log\left(1 + \frac{c}{\lambda(y)}\right) dy - c.$$

- The importance sampling intensity

$$\tilde{\lambda}(t,x) := \lambda(x) \exp\{-D\bar{W}(t,x;c^*)\} = \lambda(x) + c^*.$$

Illustration - Monte Carlo: a = 0.01, b = 5, d = 2, z = 0.25, N = 10000

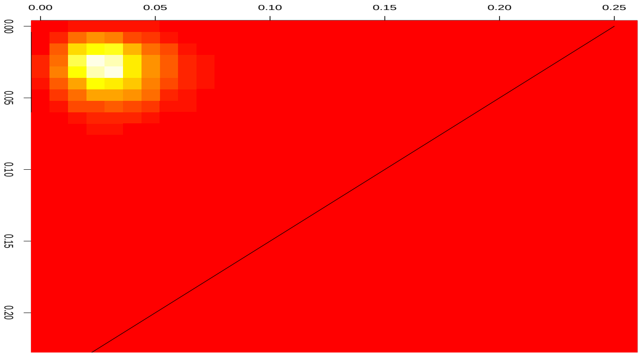

Figure : Location of the outcomes- Using Monte Carlo

Illustration - Importance sampling: a = 0.01, b = 5, d = 2, z = 0.25, N = 10000

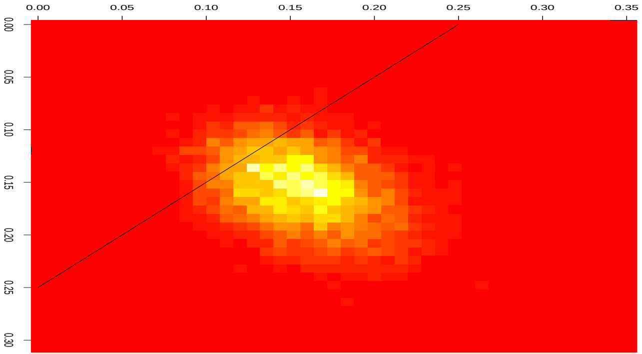

Figure : Location of outcomes- adhoc weighted empirical measure (Carmona and Crépy (2010))

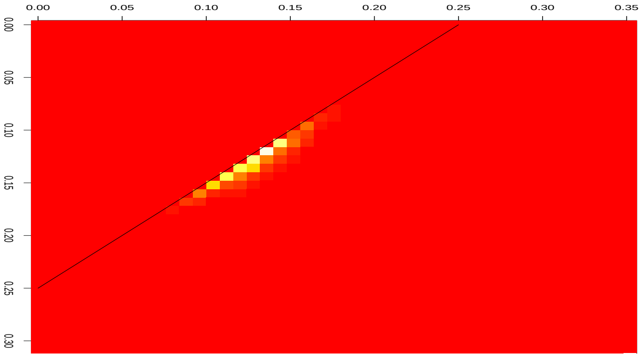

Figure : Location of outcomes- Optimally weighted empirical measure

Summary

• Optimal importance sampling can be viewed as a control problem.

• A (good) subsolution to an associated limiting PDE leads to an efficient importance sampling algorithm.

• We demonstrate how to find a good subsolutions in a credit risk model with contagion.

Main References

- Carmona, R. and Crépy, S. (2010): Importance Sampling and Interacting Particle Systems for the Estimation of Markovian Credit Portfolio Loss Distributions. *Int. J. Theor. Appl. Finan. 13(4), 577-602.*
- Dembo, A. and Zeitouni, O. (1998): Large Deviations Techniques and Applications, Springer, New York (2nd Edt.)
- Dupuis, P. and Wang, H. (2007): Subsolutions of an Issacs equation and efficient schemes for Importance Sampling. *Math. of Oper. Research 32(3), pp. 723-757.*
- Fathi, A. and Maderna, E. (2007): Weak KAM theorem on non-compact manifolds. *Nonlinear differ. equ. appli. (14), pp. 1-27.*
- Fathi, A. and Siconolfi, A. (2004): Existence of C^1 critical sub-solutions of the Hamilton-Jacobi equation. *Inventiones mathematicae, 155(2), pp.363-388.*