Calculating principal eigenfunctions of non-negative integral kernels:

particle approximations and applications

Nikolas Kantas (Imperial)

COSMOS WORKSHOP PARIS, FEB. 2-5, 201616

joint work with Nick Whiteley (Bristol)

Introduction

- On eigen-functions and related quantities of particular integral operators (e.g. un-normalised Markov transition kernels)
- Applications:
  - rare event estimation
  - stochastic control
  - statistical mechanics/physics (neutron transport, nuclear fission, particle motion etc.)

Introduction

- On eigen-functions and related quantities of particular integral operators (e.g. un-normalised Markov transition kernels)
- Applications:
  - rare event estimation
  - stochastic control
  - statistical mechanics/physics (neutron transport, nuclear fission, particle motion etc.)
- Outline:
  - Introduction on particle methods and simple Feynman-Kac models in discrete time
  - Eigen-quantities in linear algebra: finite state spaces and matrices.
  - generalisations for measurable spaces and integral operators
  - particle approximations of eigen-quantities
  - rare events estimation example

Preliminaries

- Let $M(x, dy)$ be a homogeneous Markov probability kernel on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ and let $X_0 = x$.
- **Potential** function:
$$G = e^{U}$$
where $U : \mathsf{X} \to \mathbb{R}$.
- The non-negative kernel
$$Q(x, dy) := e^{U(x)} M(x, dy)$$
defines a linear operator on functions
$$Q(\varphi)(x) := \int Q(x, dy)\, \varphi(y)$$

Some Notation

Measures: say a measure μ on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$

$$\mu(\varphi) := \int \varphi(x)\mu(dx)$$

Integral kernels:

- For K a (possibly un-normalised) Markov kernel on $\mathsf{X} \times \mathcal{B}(\mathsf{X})$,

$$K(\varphi)(x) := \int K(x, dy)\varphi(y)$$

  and

$$\mu K(\cdot) := \int \mu(dx)K(x, \cdot)$$

- the n-fold iterate of kernel K: $K^n = \underbrace{K \dots K}_{n \text{ times}}$, $K_0 = Id$

Think of kernel K being a matrix, μ a row vector and φ a vector!

Feynman-Kac models

- Following notation/style of [Del Moral 04]
- Usually in Monte Carlo we are interested to compute something like:

$$\gamma_{n,x}(\varphi) = Q^n(\varphi)(x) = \mathbb{E}_x\left[\exp\left(\sum_{k=0}^{n-1} U(X_k)\right)\varphi(X_n)\right]$$

- Simple example: particle motion in absorbing medium
  - $1 - e^{U(x)}$ probability of absorption at location x
  - $M(x, dx')$ describe Markov dynamics of neutron
  - $Q^n(1)(x)$ is probability of survival after n steps starting at $X_0 = x$.

Feynman-Kac models

- Feynman-Kac Models [Del Moral 04]:
  - more convenient to use a sequence of probability measures
  - use the following sequence

$$\eta_n(\varphi) = \frac{\mathbb{E}_x\left[\exp\left(\sum_{k=0}^{n-1} U(X_k)\right)\varphi(X_n)\right]}{\mathbb{E}_x\left[\exp\left(\sum_{k=0}^{n-1} U(X_k)\right)\right]}$$

  - Update - prediction recursion:

$$\eta_{n+1}(dy) = \frac{\eta_n Q(x, dy)}{\eta_n Q(1)}$$

Feynman-Kac models

Measure valued recursion :

$$\eta_{n+1} := \Phi(\eta_n) = \frac{\eta_n Q}{\eta_n Q(1)},$$

or in update - prediction steps

$$\eta_{n+1} = \hat{\eta}_n M, \quad \hat{\eta}_n = \frac{e^U \eta_n}{\eta_n\left(e^U\right)}$$

Product formula:

$$\gamma_{n,x}(\varphi) = \eta_n(\varphi) \prod_{k=0}^{n-1} \eta_k(e^U)$$

A particle algorithm and some approximations

A simple particle algorithm

- Initialization: sample i.i.d. $\left(\zeta_0^i\right)_{i=1}^N \sim \mu$,
- For $p = 1, ..., n$: sample i.i.d.

$$\left(\zeta_p^i\right)_{i=1}^N \Big| \left(\zeta_{p-1}^i\right)_{i=1}^N \sim \sum_{j=1}^N \frac{\exp\left[U\left(\zeta_{p-1}^j\right)\right]}{\sum_{j'=1}^N \exp\left[U\left(\zeta_{p-1}^{j'}\right)\right]} M\left(\zeta_{p-1}^j, \cdot\right)$$

Particle approximations, $n \geq 0$:

$$\eta_n^N := \frac{1}{N}\sum_{i=1}^N \delta_{\zeta_n^i}, \quad \eta_n^N(\varphi) := \frac{1}{N}\sum_{i=1}^N \varphi\left(\zeta_n^i\right),$$

$$\gamma_{n,x}^N(\varphi) := \prod_{k=0}^{n-1} \eta_k^N\left(e^U\right) \eta_n^N(\varphi)$$

The eigen problem

Under some regularity assumptions, Q has:

- an isolated, real, maximal eigen-value $\lambda_\star$,
- a positive (right) eigen-function $h_\star$,
- a positive (left) eigen-measure $\eta_\star$
- All-together

$$Q(h_\star) = \lambda_\star h_\star, \quad \text{and} \quad \eta_\star Q = \lambda_\star \eta_\star, \qquad \eta_\star(h_\star) = 1$$

- Do these quantities exist? Are they unique?
- How can they be computed?
- Is it possible to get extensions on
  - Perron-Frobenious theory
  - the power method from when $Q, h_\star, \eta_\star$ are matrix and vectors resp.?

The eigen problem - why do we care?

Applications:

- **Statistical physics**
  - $h_\star$ could be viewed as the Schrödinger ground energy state for molecules,
  - diffusion or quantum Monte Carlo
  - [Rousset 06], [R. Assaraf, M. Caffarel, & A. Khelif 00], [Makrini, B. Jourdain, and T. Lelièvre. 07].
- In **optimal control**:
  - discrete time problems with Kullback-Leibler divergence term in the stage cost
    - Q is as a multiplicative Bellman operator
    - $h_\star$ is a logarithmic transformation of the value function.
    - [Albertini & Runggaldier 88], [Todorov 08]
  - can be related to discretisations of certain continuous time models
    - [Flemming 82], [Sheu 84], [Dai Pra, Meneghini, Runggaldier,1996], [Kappen 05], [Theodorou 10]

The eigen problem - why do we care?

$h_\star$ appears in large deviations theory of Markov chains;

- [Vere-Jones 67], [Ney & Nummelin 87], [Kontoyiannis & Meyn 03]
  - Let $(X_n; n \geq 0)$ is a Markov chain with transition kernel M, $X_0 = x$, and $G(x) := e^{\alpha U(x)}$
  - Eigen-problem acts as multiplication Poisson eqn.

$$\Lambda(\alpha) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}_x \left[\exp\left(\alpha \sum_{p=0}^{n-1} U(X_p) \right) \right],$$

$$h_\star(x) = \lim_{n\to\infty} \mathbb{E}_x \left[\exp\left(\alpha \sum_{p=0}^{n-1} U(X_p) - n\Lambda(\alpha) \right) \right]$$

  - Let $\Lambda'(\alpha) = c$, $\sigma_\alpha^2 = \Lambda''(\alpha)$ and some conditions on U

$$\mathbb{P}_x \left[\sum_{p=0}^{n-1} U(X_p) > nc \right] \sim \frac{h_\star(x)}{\alpha\sqrt{2\pi\sigma_\alpha^2}} \exp\left(-n\Lambda^*(c)\right)$$

The eigen problem: the "twisted" Markov kernel

- A related object of interest:

$$P_\star(x, dx') := \frac{Q(x, dx')h_\star(x')}{h_\star(x)\lambda_\star}. \tag{1}$$

- Many applications of interest :
  - estimation of tail probabilities of Markov chains:
    - certain $P_\star$ is optimal changes of measure [Bucklew, Ney, Sadowski 90], [Dupuis & Wang 05]
  - optimal control:
    - $P_\star$ is optimally controlled Markov transition kernel.
  - particle motion in absorbing media:
    - $P_\star$ is the Markov transition kernel of a particle conditional on long-term survival [Del Moral, Miclo 03], [Del Moral & Doucet 04], [Rousset 06]
  - branching processes: [Harris 51], [Athreya 00]

Eigen -values, -measures, -functions of $Q\ (= e^U M)$

We look to compute quantities that satisfy:

$$Q(h_\star) = \lambda_\star h_\star, \quad \text{and} \quad \eta_\star Q = \lambda_\star \eta_\star, \qquad \eta_\star(h_\star) = 1$$

- In the finite state space case
  - When do these quantities exist? Are they unique?
  - How can they be computed?

Basic Perron-Frobenious theory

Let Q be a square $n \times n$ matrix with real positive entries. Then:

1. there is one real and isolated eigenvalue $\lambda_\star$ on the spectral radius and for any other eigenvalue λ, $|\lambda| < \lambda_\star$.
2. Perron vectors: for $\lambda_\star$ there exist **unique** positive right and left eigenvectors $h_\star, \eta_\star$ of size $n \times 1$ and $1 \times n$ resp. that satisfy

$$Qh_\star = \lambda_\star h_\star, \qquad \eta_\star Q = \lambda_\star \eta_\star, \qquad \eta_\star h_\star = 1$$

3. We have:

$$\lambda_\star^{-n} Q^n \to h_\star \eta_\star$$

How to compute $\lambda_\star, h_\star, \eta_\star$?

- The power method: $h_k \to h_\star$,

$$h_{k+1} = \frac{Qh_k}{\|Qh_k\|}$$

  assuming $h_0' h_\star \neq 0$. Similar for $\eta_\star$.

- Approximation for the eigenvalue $\lambda_k \to \lambda_\star$, e.g. using Rayleigh quotient:

$$\lambda_k = \arg\min_\lambda \|Qh_k - \lambda h_k\|_2 = \frac{h_k^* Q h_k}{h_k^* h_k}$$

- Gelfand's formula

$$|||Q^k|||^{1/k} \to \lambda_\star$$

General State Space: regularity conditions for Q

Recall $Q = e^U M$, M a Markov probability kernel, U a real function on X

Regularity condition (A1): there exists a probability measure ν such that

$$\frac{dQ(x,\cdot)}{d\nu}(y) = q(x,y)$$

and

$$0 < \epsilon^- \leq q(x,y) \leq \epsilon^+ < \infty$$

- Under (A1) Q is aperiodic and irreducible.
- A bit restrictive but hopefully this can be relaxed, e.g. [Kontoyiannis and Meyn 03] etc.

Perron-Frobenius Theory for Q

- [...,Nummelin 84] Under (A1) there exist: a maximal and isolated eigenvalue $\lambda_\star$, a positive eigen-function, probability eigen-measure:

$$Q(h_\star) = \lambda_\star h_\star, \quad \text{and} \quad \eta_\star Q = \lambda_\star \eta_\star, \qquad \eta_\star(h_\star) = 1$$

and these are unique.

Perron-Frobenius Theory for Q

- [...,Nummelin 84] Under (A1) there exist: a maximal and isolated eigenvalue $\lambda_\star$, a positive eigen-function, probability eigen-measure:

$$Q\,(h_\star) = \lambda_\star h_\star, \quad \text{and} \quad \eta_\star Q = \lambda_\star \eta_\star, \qquad \eta_\star\,(h_\star) = 1$$

  and these are unique.

- Multiplicative Ergodic Theorem: For any $n \geq 1$,

$$\sup_{x \in \mathsf{X}} \sup_{|\varphi|<1} \left| \lambda_\star^{-n} Q^n\,(\varphi)\,(x) - h_\star(x)\eta_\star(\varphi) \right| \;\leq\; 2\left(\frac{\epsilon^+}{\epsilon^-}\right)^2 \rho^n$$

  or

$$|||\lambda_\star^{-n} Q_\star^n - h_\star \otimes \eta_\star||| \leq 2\rho^n \left(\frac{\epsilon^+}{\epsilon^-}\right)^2$$

  where $\rho = 1 - (\epsilon^-/\epsilon^+)$.

The twisted probability kernel (Doob's h kernel)

- The MET is derived using the properties of the twisted Markov kernel
$$P_\star(x, dy) := \frac{Q(x, dy)h_\star(y)}{h_\star(x)\lambda_\star}. \tag{2}$$
- $P_\star$ is ergodic with a unique invariant probability distribution, denoted by $\pi_\star$ and for all $n \geq 1$,
$$\sup_{x \in \mathsf{X}} \sup_{A \in \mathcal{B}(\mathsf{X})} |P_\star^n(x, A) - \pi_\star(A)| \leq 2\rho^n \tag{3}$$
$$d\pi_\star / d\eta_\star = h_\star \tag{4}$$
or
$$|||P_\star^n - 1 \otimes \pi_\star||| \leq 2\rho^n$$
where $\rho = 1 - (\epsilon^- / \epsilon^+)$.

How to approximate $\eta_\star$,$\lambda_\star$?

- [Del Moral, Miclo 02], [Del Moral, Doucet 04], [Rousset 06]
- Consider the forward recursion for $n \geq 0$:

$$\begin{aligned}
\eta_{n+1} &= \Phi(\eta_n) = \frac{\eta_n Q}{\eta_n Q(1)}, \\
\lambda_n &= \eta_n Q(1) = \eta_n(e^U)
\end{aligned}$$

- Recall

$$\begin{aligned}
\eta_\star Q = \lambda_\star \eta_\star \quad &\Rightarrow \quad \eta_\star Q(1) = \eta_\star(e^U) = \lambda_\star. \\
&\Rightarrow \quad \Phi\left(\eta_\star\right) = \eta_\star
\end{aligned}$$

- In fact, for any initial condition η_0

$$\begin{aligned}
\|\eta_{n+1} - \eta_\star\| &\leq C\rho^n \\
|\lambda_n - \lambda_\star| &\leq C'\rho^n
\end{aligned}$$

How to approximate $h_\star$, $P_\star$?

- Often this is enough
  - when M is μ- reversible then $\frac{h_\star}{\mu(h_\star)} = \frac{d\eta_\star}{d\mu}$
  - [Rousset 06], [Makrini, B. Jourdain, and T. Lelièvre. 07].

How to approximate $h_\star$, $P_\star$?

- Often this is enough
  - when M is μ- reversible then $\frac{h_\star}{\mu(h_\star)} = \frac{d\eta_\star}{d\mu}$
  - [Rousset 06], [Makrini, B. Jourdain, and T. Lelièvre. 07].

- When this does not apply but still within (A1)
  - Keep previous forward sequence $(\eta_p, \lambda_p)_{p\geq 0}$
  - Find a sequence of functions $(h_p)_{p\geq 0}$ such that

$$Q\left(h_{p+1}\right) = \lambda_p h_p, \quad \eta_p Q = \lambda_{p+1}\eta_{p+1}, \quad \eta_{p+1}(h_{p+1}) = 1$$

How to approximate $h_\star$, $P_\star$?

- Backward recursion
- Set $h_{n,n}(x) := 1$, for $0 \leq p < n$.

$$h_{p,n}(x) := \frac{Q\left(h_{p+1,n}\right)}{\lambda_p} = \frac{Q^{(n-p)}(1)(x)}{\prod_{\ell=p}^{n-1} \lambda_\ell},$$

and

$$P_{p,n}(x, dx') := \frac{Q(x, dx') h_{p,n}}{\lambda_{p-1} h_{p-1}}$$

How to approximate $h_\star$, $P_\star$?

- Backward recursion
- Set $h_{n,n}(x) := 1$, for $0 \leq p < n$.

$$h_{p,n}(x) := \frac{Q\left(h_{p+1,n}\right)}{\lambda_p} = \frac{Q^{(n-p)}(1)(x)}{\prod_{\ell=p}^{n-1} \lambda_\ell},$$

and

$$P_{p,n}(x, dx') := \frac{Q(x, dx')h_{p,n}}{\lambda_{p-1} h_{p-1}}$$

- We can invoke the MET

$$|||\lambda_\star^{-n} Q_\star^{(n)} - h_\star \otimes \eta_\star||| \leq 2\rho^n \left(\frac{\epsilon^+}{\epsilon^-}\right)^2$$

to deduce

$$\begin{aligned} \|h_{p,n} - h_\star\| &\leq C\rho^{p\wedge(n-p)} \\ |||P_{p,n} - P_\star||| &\leq C'\rho^{p\wedge(n-p)} \end{aligned}$$

An ideal algorithm like the power method

<u>Forward</u>

Initialization: Set $\eta_0 = \mu$,

For $p = 1, \ldots, 2n$, :

Set $\eta_p = \frac{\eta_n Q}{\lambda_n}, \quad \lambda_n = \eta_n Q(1)$

<u>Backward</u>

Initialization: Set $h_{2n,2n}(x) = 1, \quad x \in \mathsf{X}$

For $p = 2n-1, \ldots, n$, :

Set $h_{p,2n}(x) = (\lambda_p)^{-1} Q\left(h_{p+1,2n}\right)(x), \quad x \in \mathsf{X}$

Set $P_{p+1,n}(x, dx') := \frac{Q(x,dx')h_{p+1,n}(x')}{\lambda_p h_p(x)}, \quad x, x' \in \mathsf{X}$

$$\begin{aligned} \|h_{n,2n} - h_\star\| &\leq C\rho^n \\ |||P_{p,n} - P_\star||| &\leq C'\rho^n \end{aligned}$$

A particle algorithm

- Forward
  - Initialization: sample i.i.d. $\left(\zeta_0^i\right)_{i=1}^N \sim \mu$,
  - For $p = 1, ..., 2n$: sample i.i.d.

$$\left(\zeta_p^i\right)_{i=1}^N \Big| \left(\zeta_{p-1}^i\right)_{i=1}^N \sim \sum_{j=1}^N \frac{\exp\left[U\left(\zeta_{p-1}^j\right)\right]}{\sum_{j'=1}^N \exp\left[U\left(\zeta_{p-1}^{j'}\right)\right]} M\left(\zeta_{p-1}^j, \cdot\right),$$

A particle algorithm

- Forward
  - Initialization: sample i.i.d. $(\zeta_0^i)_{i=1}^N \sim \mu$,
  - For $p = 1, ..., 2n$: sample i.i.d.

$$\left(\zeta_p^i\right)_{i=1}^N \Big| \left(\zeta_{p-1}^i\right)_{i=1}^N \sim \sum_{j=1}^N \frac{\exp\left[U\left(\zeta_{p-1}^j\right)\right]}{\sum_{j'=1}^N \exp\left[U\left(\zeta_{p-1}^{j'}\right)\right]} M\left(\zeta_{p-1}^j, \cdot\right),$$

- Backward
  - Initialization: set $h_{2n,2n}(x) = 1, \quad x \in \mathsf{X}$
  - For $p = 2n-1, ..., n$, set

$$h_{p,2n}^N(x) = \sum_{j=1}^N \frac{q(x, \zeta_{p+1}^j)}{\sum_{i=1}^N q(\zeta_p^i, \zeta_{p+1}^j)} h_{p+1,2n}^N(\zeta_{p+1}^j), \quad x \in \mathsf{X}$$

Particle Approximations

- (Forward pass) Target eigen-measure and eigen-value

$$\eta_n^N := \frac{1}{N} \sum_{i=1}^{N} \delta_{\zeta_n^i}, \quad \eta_n^N(\varphi) := \frac{1}{N} \sum_{i=1}^{N} \varphi\left(\zeta_n^i\right), \quad \lambda_n^N := \eta_k^N\left(e^U\right)$$

Particle Approximations

- (Forward pass) Target eigen-measure and eigen-value

$$\eta_n^N := \frac{1}{N}\sum_{i=1}^{N}\delta_{\zeta_n^i}, \quad \eta_n^N(\varphi) := \frac{1}{N}\sum_{i=1}^{N}\varphi\left(\zeta_n^i\right), \quad \lambda_n^N := \eta_k^N\left(e^U\right)$$

- (Backward pass) eigen-function

$$h_{n,2n}^N(x) = \sum_{j=1}^{N}\frac{q(x,\zeta_{n+1}^j)}{\sum_{i=1}^{N} q(\zeta_n^i,\zeta_{n+1}^j)} h_{n+1,2n}^N(\zeta_{n+1}^j)$$

- (Backward pass) twisted kernel:

$$P_{n,2n}^N(x,dx') := \frac{1}{h_{n-1,2n}^N(x)}\sum_{j=1}^{N}\frac{q(x,\zeta_n^j)}{\sum_{i=1}^{N} q(\zeta_{n-1}^i,\zeta_n^j)} h_{n,2n}^N(\zeta_n^j)\delta_{\zeta_n^j}(dx').$$

L_r error estimates

Under (A1) we have $\forall n \geq 1, N \geq 1, r \geq 1$ there exist constants $B(r), C$ such that:

$$\sup_{x \in \mathsf{X}} \mathbb{E}_N \left[\left| h^N_{n,2n}(x) - h_\star(x) \right|^r \right]^{1/r} \leq \frac{B_h(r)}{\sqrt{N}} + C_h \rho^n \quad (5)$$

$$\sup_{x \in \mathsf{X}} \sup_{A \in \mathcal{B}(\mathsf{X})} \mathbb{E}_N \left[\left| P^N_{n,2n}(x, A) - P_\star(x, A) \right|^r \right]^{1/r} \leq \frac{B_P(r)}{\sqrt{N}} + C_P \rho^n, \quad (6)$$

where $\rho = (1 - (\epsilon^- / \epsilon^+))$.

In part appeal to earlier results for particle smoothers [Del Moral, Doucet, Singh 10] and [Douc, Garivier, Moulines & Olsson 11]

An application on rare events estimation

- Let $(X_n; n \geq 0)$ be a Markov chain with transition M initialized from some $X_0 = x$.
- Also let $U : \mathsf{X} \to [-1, 1]$
- Our objective is for some $\delta \in (0, 1)$ and $m \geq 1$, to estimate the deviation probability

$$\pi_m(\delta) := \mathbb{P}_x \left(\sum_{p=1}^{m} U(X_p) > m\delta \right). \tag{7}$$

where $\mathbb{P}_x$ is the law of the chain.

Sequential Importance Sampling (IS) for rare events

- Let $\mathcal{C}$ be the collection of Markov transitions $\overline{M}$: there exist $0 < \bar{\epsilon}^{-}, \bar{\epsilon}^{+} < \infty$ and a probability measure $\bar{\nu}$ such that $\nu \ll \bar{\nu}$

$$(\mathcal{C}) \quad \bar{\nu}(\cdot)\,\bar{\epsilon}^{-} \leq \overline{M}(x,\cdot) \leq \bar{\epsilon}^{+}\bar{\nu}(\cdot)\,,\ \forall x, \quad \int \left(\frac{d\nu}{d\bar{\nu}}(x)\right)^2 \bar{\nu}(dx) < \infty$$

Sequential Importance Sampling (IS) for rare events

- Let $\mathcal{C}$ be the collection of Markov transitions $\overline{M}$: there exist $0 < \bar{\epsilon}^-, \bar{\epsilon}^+ < \infty$ and a probability measure $\bar{\nu}$ such that $\nu \ll \bar{\nu}$

$$(\mathcal{C}) \quad \bar{\nu}(\cdot)\,\bar{\epsilon}^- \leq \overline{M}(x,\cdot) \leq \bar{\epsilon}^+ \bar{\nu}(\cdot)\,,\ \forall x, \quad \int \left(\frac{d\nu}{d\bar{\nu}}(x)\right)^2 \bar{\nu}(dx) < \infty$$

- The sequential importance sampling (IS) estimator of $\pi_m(\delta)$ is

$$\widehat{\pi}_m(\delta, L) := \frac{1}{L}\sum_{i=1}^{L} \mathbb{I}\left[\sum_{p=1}^{m} U\left(X_p^i\right) > m\delta\right] \frac{d\mathbb{P}_x}{d\overline{\mathbb{P}}_x}\left(X_0^i, \ldots, X_m^i\right), \tag{8}$$

where $\left\{\left(X_0^i, X_1^i, \ldots,\right); i = 1, \ldots, L\right\}$ are L iid Markov chains, each with transition $\overline{M}$ and law denoted by $\overline{\mathbb{P}}_x$

Importance Sampling for rare events: twisting the kernel

- AIM: the relative variance of IS

$$\overline{\mathbb{E}}_x \left[\left(\frac{\widehat{\pi}_m \left(\delta, L\right)}{\pi_m(\delta)} - 1 \right)^2 \right] = \frac{1}{L} \left(\frac{\overline{\mathbb{E}}_x \left[\widehat{\pi}_m \left(\delta, 1\right)^2 \right]}{\pi_m(\delta)^2} - 1 \right) \qquad (9)$$

not to grow exponentially fast in m.

Importance Sampling for rare events: twisting the kernel

- AIM: the relative variance of IS

$$\overline{\mathbb{E}}_x\left[\left(\frac{\widehat{\pi}_m(\delta, L)}{\pi_m(\delta)} - 1\right)^2\right] = \frac{1}{L}\left(\frac{\overline{\mathbb{E}}_x\left[\widehat{\pi}_m(\delta, 1)^2\right]}{\pi_m(\delta)^2} - 1\right) \qquad (9)$$

  not to grow exponentially fast in m.
- Consider

$$G_{\alpha'}(x) := e^{\alpha' U(x)}, \qquad Q_{\alpha'}(x, dx') := G_{\alpha'}(x) M(x, dx')$$

  and

$$\begin{aligned} \Lambda_\star(\alpha) &:= \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}_x\left[\exp\left(\alpha \sum_{p=0}^{n-1} U(X_p)\right)\right], \quad \alpha \in \mathbb{R} \\ I(t) &:= \sup_{\alpha\in\mathbb{R}} \left[t\alpha - \Lambda_\star(\alpha)\right], \quad t \in \mathbb{R} \end{aligned}$$

Importance Sampling with large deviations Bucklew et al. 90

1. $I(t)$ is a non-negative, strictly convex function with $I(t) = 0$ if and only if $t = \Lambda'_\star(0)$.
2. For any $\delta \in (0,1)$, the following large deviation principle holds

$$\lim_{m\to\infty} \frac{1}{m} \log \pi_m(\delta) = - \inf_{t\in[\delta,\infty)} I(t).$$

3. For any $\delta \in (0,1)$ and $\overline{M}$ in $\mathcal{C}$, the importance sampling estimator satisfies

$$\lim_{m\to\infty} \frac{1}{m} \log \overline{\mathbb{E}}_x \left[\widehat{\pi}_m(\delta,1)^2\right] \geq -2 \inf_{t\in[\delta,\infty)} I(t). \tag{10}$$

4. For any $\delta \in (0,1)$ and α the unique solution of

$$\Lambda'_\star(\alpha) = \delta$$

the twisted kernel $P^\alpha_\star$ is the unique member of $\mathcal{C}$ for which equality holds (*asymptotically efficient.*)

Importance Sampling for rare events using lack of bias

- Let $\left(\check{X}_p; p = 0, ..., m\right)$ be a non-homogeneous Markov chain with transitions

$$\check{X}_0 = x, \qquad \check{X}_p \sim P^N_{\alpha,n+p,2n}\left(\check{X}_{p-1}, \cdot\right), \quad p \geq 1.$$

- Then we have the following lack of bias property:

$$\mathbb{E}_N\left[\mathbb{I}\left[\sum_{p=1}^{m} U\left(\check{X}_p\right) > m\delta\right] \prod_{p=0}^{m-1} \frac{\lambda^N_{\alpha,n+p}}{G_\alpha\left(\check{X}_p\right)} \frac{h^N_{\alpha,n,2n}\left(\check{X}_0\right)}{h^N_{\alpha,n+m,2n}\left(\check{X}_m\right)}\right] = \pi_m(\delta),$$

$\check{\mathbb{E}}_{N,\alpha}$ for expectation w.r.t. the joint law of $\left(\check{X}_p\right)$ and the particle system.

Numerical example

Take $\mathsf{X} = [-c, c]$ and consider an ergodic Gaussian transition kernel with support restricted to $[-c, c]$,

$$M(x, dy) = \frac{\exp\left(-\frac{1}{2}\left(y - \frac{x}{2}\right)^2\right)}{\left(1 - \operatorname{erf}\left(\frac{c + x/2}{\sqrt{2}}\right)\right)\sqrt{2\pi}} \mathbb{I}_{[-c,c]}(y)dy,$$

Consider U defined by

$$U(x) = \begin{cases} -1 & x \leq -1 \\ x & x \in (-1, 1) \\ 1 & x \geq 1. \end{cases}$$

Numerical example

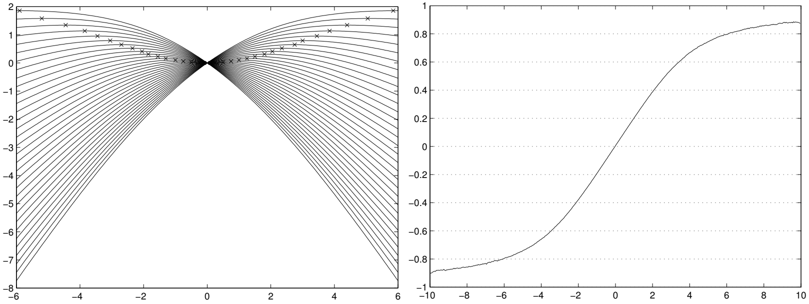

Figure : Left: each of the solid curves shows an approximation of $[\alpha t - \Lambda_\star(\alpha)]$ against α, with each curve corresponding to a different value of t in the range $[-0.8, 0.8]$. The cross on each curve indicates its maximum and thus approximates $\sup_\alpha [\alpha t - \Lambda_\star(\alpha)] = I(t)$. Right: $\Lambda'_\star(\alpha)$ against α approximated using finite differences.

Numerical example

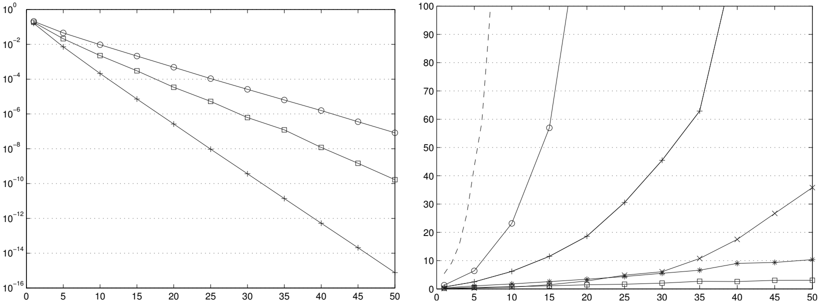

Figure : Left: estimated value of $\pi_m(\delta)$ against m, for: $\circ$,$\delta = 0.8$; $\square$,$\delta = 0.9$, and $+$, $\delta = 0.99$. Right: solid lines show sample relative variance of the estimated value of $\pi_m(0.9)$ against m for: $\circ$, $\alpha = 1$; $+$,$\alpha = 2$; $*$, $\alpha = 4$; $\square$, $\alpha = 8$; and $\times$,$\alpha = 16$. Dashed line shows sample relative variance of $\widehat{\pi}_m(0.9, 1)$ in the case $\overline{M} = M$.

Conclusions/Extensions

- Some interesting examples/extensions on stochastic control:
  - $h_p, h_\star$ are value functions for particular finite/ infinite horizon problems resp.
- In the rare events example, can use also other ideas from SMC or adaptive IS [Dupuis and Wang 05] to improve things.
- When $h_p(x), h_\star(x)$ are the inference objective
  - there is some interest in variance reduction
- This is a batch scheme with cost $O(N^2 n)$
  - can this be reduced in some way?

References

- J.A. Bucklew, P. Ney, J.S. Sadowsky (1990) Monte Carlo simulations and large deviations theory for uniformly recurrent Markov chains J. Appl. Prob.
- P. Del Moral and L. Miclo. (2003) Particle approximations of Lyapunov exponents connected to Schrodinger operators and Feynman-Kac semigroups. ESAIM Probab. Stat.
- P. Del Moral, A. Doucet, and S.S. Singh. (2011) A backward particle interpretation of Feynman-Kac formulae. ESAIM Math. Model. Numer. Anal.
- R. Douc, Garivier, A. and Moulines, E. and Olsson, J. (2011) Sequential Monte Carlo smoothing for general state space hidden Markov models. Ann. Appl. Probab..
- E. Nummelin. (1984) General irreducible Markov chains and non-negative operators. CUP.
- I. Kontoyiannis and S. P. Meyn (2003) Spectral theory and limit theorems for geometrically ergodic Markov processes,Ann. Appl. Probab. 13 (1), 304-362.
- N. Whiteley and N. Kantas (2012). A particle method for approximating principal eigen-functions and related quantities. Submitted.
- N. Whiteley, N. Kantas and A. Jasra, (2012) Linear Variance Bounds for Particle Approximations of Time-Homogeneous Feynman-Kac Formulae, SPA.
- N. Whiteley (2013) Stability properties of some particle filters. Ann. Appl. Probab...

Appendix: relaxing (A1)

- Similar to [Whiteley, N.K., Jasra 12], [Whiteley 13]
- Use assumptions from [Kontoyiannis & Meyn 03]:
  - using multiplicative drift conditions , i.e. there exists $b_d < \infty$ and Lyapunov function V such that:

$$Q\left(e^{V}\right) \quad \leq \quad e^{V(1-\delta)+b_d \mathbb{I}_{C_d}}.$$

  - a stronger MET is derived for Q
  - can be verified in general for following cases:

  1. bounded functions U and non-ergodic kernels M
  2. unbounded above functions U and multiplicative ergodic kernels M

- Can be verified in practice for realistic examples

Optimal Control

Let $(X_n; n \geq 0)$ be a controlled Markov chain initialized from $X_0 = x$ and $X_n \sim M^{f_{n-1}}(X_{n-1}, \cdot)$

$$V_0(x) = \inf_{f \in \mathcal{H}^n} \mathbb{E}^f_{x,0} \left[\sum_{p=0}^{n-1} \left(U(X_p) + \mathcal{KL}\left(\check{M}^{f_p} \middle\| M \right)(X_p) \right) + \Omega(X_n) \right],$$

$$V_\star(x) = \inf_{f \in \mathcal{H}^{\mathbb{N}}} \limsup_{n \to \infty} \frac{1}{n} \mathbb{E}^f_{x,0} \left[\sum_{p=0}^{n} \left(U(X_p) + \mathcal{KL}\left(M^{f_p} \middle\| M \right)(X_p) \right) \right]$$

Finite horizon setup

- Let $V_n = \Omega$, $Q = e^{-U}M$, $\lambda_p = e^{-\Lambda_p}$ and $V_{p+1} = -\log h_p$ the backward recursion

$$Q\left(h_{n+1}\right) = \lambda_n h_n$$

corresponds to Bellman eqn.

$$V_p(x) = U(x) - \Lambda_p + \inf_{f_p \in \mathcal{H}} \left\{ \mathcal{KL}\left(M^{f_p} \middle\| M \right)(x) + M^{f_p}\left(V_{p+1}\right)(x) \right\}$$

- For $p = 1, \ldots, n$
  - Run forward particle system to compute particle approximations for (η_p),
- For $p = n, n-1 \ldots, 1$
  - Run backwards in time to compute particle approximations for h_p and P_p

Infinite horizon problem: particle value iteration

- Similarly, $V_\star(x) = -\log h_\star(x)$, $\varsigma_\star = -\log \lambda_\star$

$$Q\left(h_\star\right) = \lambda_\star h_\star$$

is equivalent to a solution $(V_\star, \Lambda_\star)$ of the average-cost optimality equation:

$$V_\star(x) + \Lambda_\star = \inf_{h \in \mathcal{H}} \left[U(x) + \mathrm{KL}\left(\check{M}^h \middle\| M \right)(x) + \check{M}^h\left(V_\star\right)(x) \right].$$

- Use previous algorithm for n very large
  - forward particle system up to time $2n$ to compute particle approximations (η_p^N),
  - then backwards to time n to compute particle approximations $h_{n,2n}^N$ and $P_{n,2n}^N$

Example: Gaussian model

Consider the controlled dynamics:

$$X_n = \begin{bmatrix} X_n^1 \\ V_n^1 \end{bmatrix} = \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix} X_{n-1} + \begin{bmatrix} \tau & \tau^2 \\ 0 & \tau \end{bmatrix} (W_n + A_n)$$

consider the state-dependent-only part of the stage cost:

$$U(x) \propto (1 - \mathbb{I}_{(-\delta,\delta)}(x^1))$$

which penalises states outside $(-\delta, \delta)$.

Example: Gaussian model

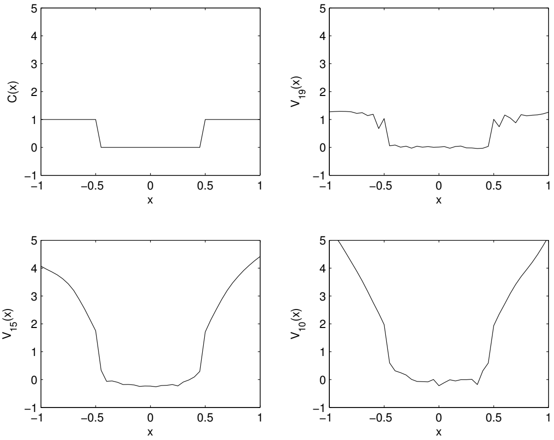

Figure : Estimated horizon average cost optimal value function $V_n(x)$ against x for various n. Here $T = 20$, $\Psi_T = C$, $\tau = 1$, $N = 500$

Example: Cox-Ingersoll-Ross process

- Control free model for M is Euler discretisation of CIR

$$dX_t = \theta\,(\mu - X_t)\,dt + \sigma\sqrt{X_t}dW_t,$$

  where $\{W_t\}$ is standard 1-D Brownian motion, $\theta > 0$ is the reversion rate, $\mu > 0$ is the level of mean reversion and $\sigma > 0$ specifies the volatility.
- Stage cost specified by

$$U(x) = 2\mathbb{I}_{[0,10-\delta]}(x) + \mathbb{I}_{[10+\delta,\infty)}(x), \tag{11}$$

  which penalises states outside $(10-\delta, 10+\delta)$.

Example: Cox-Ingersoll-Ross process

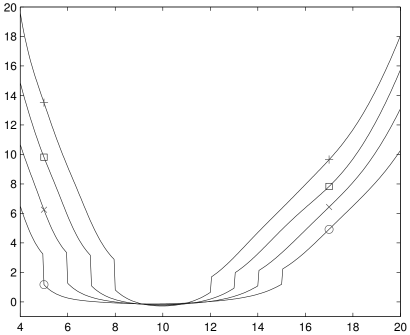

Figure : Estimated infinite-horizon average cost optimal value function $V_\star(x)$ against x for: $\circ, \delta = 5; \times, \delta = 4; \Box, \delta = 3; +, \delta = 2$.

Appendix : connection to some pde's

Numerical solutions for some class of parabolic PDEs for $t \in [0, T]$

$$(v)_t + a\,(v)_x + \frac{1}{2}\sigma^2\,(v)_{xx} + \varphi v = 0, \quad v(\cdot, T) = \psi(\cdot)$$

using $Q^{(n-p)}(\psi)(x) =$

$$h_{p,n}(x) \prod_{l=p}^{n-1} \lambda_l \approx \mathbb{E}_{tx}\left[\exp\left(\int_t^T U(X_s)ds\right)\varphi(X_T)\right] = v(x,t)$$

where the expectation is taken conditional to $X_t = x$ w.r.t.

$$dX_t = a(X_t)dt + \sigma(X_t)dW_t.$$

Appendix: a stochastic control example

- For the following controlled Markov chain

$$dX_t = (a(X_t) + B(X_t)A_t)\,dt + \sigma(X_t)dW_t,$$

let the total cost or value function for $t \geq 0$:

$$V(x,t) = \inf_{U \in L^2(\cdot), s \in (t,T)} \mathbb{E}_{tx}\left[\int_t^T L(X_s, A_s)ds + \Psi(X_T)\right],$$

with the stage cost being:

$$L(x,A) = C(x) + \frac{1}{2}A' R(x)A,$$

Appendix: a stochastic control example

- For the following controlled Markov chain

$$dX_t = (a(X_t) + B(X_t)A_t)\,dt + \sigma(X_t)dW_t,$$

let the total cost or value function for $t \geq 0$:

$$V(x,t) = \inf_{U \in L^2(\cdot), s \in (t,T)} \mathbb{E}_{tx}\left[\int_t^T L(X_s, A_s)ds + \Psi(X_T)\right],$$

with the stage cost being:

$$L(x,A) = C(x) + \frac{1}{2}A^{'}R(x)A,$$

- IF we have $BRB' = \sigma\sigma'$ then the Hamilton-Jacobi-Bellman equation implies

$$V(x,t) = -\log v(x,t), \quad C = -U, \quad \Psi = -\log\varphi \qquad (12)$$

with $v(x,t)$ as in the previous slide and for the optimal control:

$$A^*(x,t) = R(x)^{-1}B(x)'\nabla_x v(x,t).$$