Sequential Monte Carlo for Risk Sensitive Control

N. Kantas[1] A. Doucet[2] S.S. Singh[1]
1 Cambridge University Engineering Dept., Cambridge CB2 1PZ, UK
2 The Institute of Statistical Mathematics, Tokyo 106-8569, Japan

Greek Stochastics α, Lefkada, 31 August 2009

Problem statement

- **Model**: Let $\{X_n\}_{n\geq 0}$ be a $\mathcal{X}$ $(\subseteq \mathbb{R}^{n_x})$ -valued Markov process defined on a (measurable) space $(\Omega, \mathcal{F})$.

$$X_0 \sim \nu(\cdot), \quad X_n \sim M_\theta(X_{n-1}, \cdot), \tag{1}$$

where for the parameter we assume $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$, Θ is open.

- **Objective**: Estimate θ^* such that

$$\theta^* = \arg\min_{\theta\in\Theta} J_\beta(\theta),$$

with

$$J_\beta(\theta) = \lim\sup_{n\to\infty} \frac{1}{\beta n} \log\left(\mathbb{E}\left[\exp \sum_{p=1}^{n} \beta V_\theta(X_p)\right]\right).$$

Outline

- Introduce Risk Sensitive Markov Decision Processes
- Pose the problem a sequence of Feynman Kac (F-K) distributions
- Show direct analogy with Recursive Maximum Likelihood
- Use Sequential Monte Carlo (SMC) to compute the optimal policy

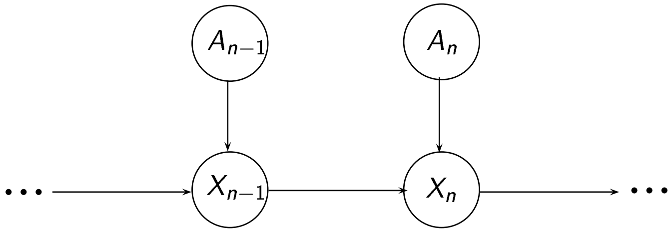
Introduction: Markov Decision Process

- Let $\{X_n\}_{n\geq 0}$ be a Markov process depending on some a $\mathcal{A}$ ($\subseteq \mathbb{R}^{n_a}$) -valued action sequence $\{A_n\}_{n\geq 0}$
  - with initial distribution μ
  - a family of transition kernels $\{M_n\}_{n\geq 0}$ such that

$$\mathbb{P}(X_n \in dx_n | X_{0:n-1} = x_{0:n-1}, A_{1:n} = a_{1:n}) = M(x_{n-1}, a_n, dx_n).$$

- Policy ζ: the sequence of mappings $\{\Pi_n\}_{n\geq 0}$.
  - Randomised:Π_n is a kernel with domain $\mathcal{X} \times \mathcal{A} \to \mathcal{P}(\mathcal{A})$
  - Deterministic: map $\mathcal{X} \to \mathcal{A}$, using $A_n = \Pi_n(X_n)$

Introduction: Markov Decision Process



Introduction: Risk Sensitive Cost

- Infinite horizon risk sensitive cost (e.g. Di Masi-Stettner [3])

$$J(\zeta) = \limsup_{n\to\infty} \frac{1}{\beta n} \log \left(\mathbb{E}_{x_0} \left[\exp \sum_{p=1}^{n} \beta V(X_p, \Pi_p(X_p)) \right] \right),$$

  where β is a risk constant.
- Problem find a policy ζ^* such that

$$\zeta^* = \arg\inf_{\zeta} J(\zeta).$$

  - For $\beta < 0$, *risk averse.*
  - For $\beta > 0$, *risk preferring.*
  - For $\beta \to 0$, *risk neutral*, i.e. we minimise the infinite horizon average cost

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}_{x_0} \left[V(X_k, \Pi_k(X_k)) \right].$$

Introduction: Risk Sensitive Cost

- Assume we can parameterise the policy through a parameter $\theta \in \Theta$.
- Express:
  - the state's transition density as $M_\theta(x_{n-1}, x_n)$,
  - the instantaneous cost $V(X_n, A_n)$ as $V(X_n, \Pi_\theta(X_n))$ or more simply as $V_\theta(X_n)$.
- Seek to minimise

$$J(\theta) = \limsup_{n\to\infty} \frac{1}{\beta n} \log\left(\mathbb{E}_{x_0}\left[\exp \sum_{p=1}^{n} \beta V_\theta(X_p)\right]\right).$$

Toy Example: Linear Gaussian Quadratic Regulator (LQR)

- Linear Gaussian State Space Model

$$X_n = HX_{n-1} + A_n + \sigma V_n,$$

where $X_0 \sim \mathcal{N}(0,1)$ and $V_n \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$.

- Instantaneous Quadratic Cost:

$$V(X_n, A_n) = \frac{1}{2} X_n^T Q X_n + \frac{1}{2} A_n^T R A_n.$$

- State Feedback Policy, (Whittle [6]):

$$A_n = \theta X_n.$$

Example: Risk Sensitive LQR

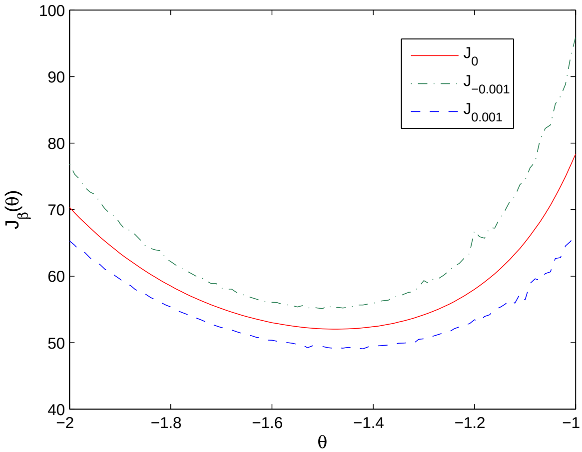


Figure: Plot $J_\beta(\theta)$ against θ for the cases when $\beta = \{-0.001, 0, 0.001\}$.

Method outline

- Formulate the problem as a Feynman Kac model (Del Moral [1])
- Compute SMC approximations of the flow and gradients
- Compute estimates of θ^*using gradients

Feynman Kac model

Consider the F-K models for the pair (G_θ, M_θ):

$$\text{Prediction: } \eta_n(dx) = \int \mu_{n-1}(dx') M_\theta(x', dx),$$

$$\text{Update: } \mu_n(dx) = \frac{1}{Z_n} G_\theta(x) \eta_n(dx),$$

where $\mu_0 = \nu$ and

$$Z_n = \int \prod_{p=1}^{n} G_\theta(x_p) M_\theta(x_{p-1}, dx_p) \nu(dx_0).$$

Note that

$$\eta_n(G_\theta) = \int \mu_{n-1}(dx') M_\theta(x', dx) G_\theta(x) = \frac{Z_n}{Z_{n-1}},$$

$$Z_n = \prod_{p=1}^{n} \eta_p(G_\theta) \qquad (2)$$

Some Assumptions

- **(A1)** Measurability,[1]. For any $x' \in \mathcal{X}$, the pairs (G_θ, M) satisfy

$$M(G_\theta) = \int G_\theta(x) M_\theta(x', dx) > 0,$$

$$\sup_{x' \in \mathcal{X}} |M(G_\theta)(x')| < \infty.$$

- **(A2)** Strong Mixing Conditions, [1, 5]. There exists a probability measure κ on $\mathcal{X}$, positive for all values of $x \in \mathcal{X}$, and constants $0 < \lambda, g_-, g_+ < \infty$ such that for all , $(x, x') \in \mathcal{X} \times \mathcal{X}$

$$\frac{1}{\lambda}\kappa(x') \leq M(x, x') \leq \lambda\kappa(x')$$

$$g_- \leq G_\theta(x) \leq g_+$$

Define F-K Potential for Risk Sensitive MDP

- Define potential function as the unnormalised Boltzmann-Gibbs measure

$$G_\theta(X_n) = \exp(\beta V_\theta(X_n))$$

Therefore,

$$J(\theta) = \beta^{-1} \limsup_{n\to\infty} \frac{1}{n} \log \left(\int \prod_{p=1}^{n} G_\theta(x_p) M_\theta(x_{p-1}, dx_p) \nu(dx_0) \right).$$

$$= \beta^{-1} \limsup_{n\to\infty} \frac{1}{n} \log \left(\prod_{p=1}^{n} \eta_p(G_\theta) \right)$$

$$= \beta^{-1} \limsup_{n\to\infty} \frac{1}{n} \sum_{p=1}^{n} \log \left(\eta_p(G_\theta) \right)$$

Similarity with Maximum Likelihood

- In Hidden Markov models we observe only:

$$Y_n | \left(X_{0:n} = x_{0:n}, Y_{0:T} = y_{0:T}\right) \sim g_\theta(\cdot|x_n)$$

- If we set $G_\theta(x) = g_\theta(y_n|x_n)$ we would be interested in

$$\begin{aligned} J(\theta) &= \lim_{n\to\infty} \frac{1}{n} \sum_{p=1}^{n} \log\left(\eta_p(G_\theta)\right) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{p=1}^{n} \log p(Y_p | Y_{0:p-1}) \\ &= \lim_{n\to\infty} \frac{1}{n} \log p(Y_{0:p}) \end{aligned}$$

  i.e. in long term average log-likelihood.
- In this case θ^*can be estimated **on-line** using Recursive Maximum Likelihood (RML)

Stochastic Approximation

- Under ergodicity and regularity assumptions (e.g. A2), based on Del Moral and Doucet [2]:

$$\mu_n \overset{n\to\infty}{\longrightarrow} \mu_\infty$$

$$\eta_n \overset{n\to\infty}{\longrightarrow} \eta_\infty$$

$$\frac{1}{n}\sum_{p=1}^{n} \log\left(\eta_n(G_\theta)\right) \overset{n\to\infty}{\longrightarrow} \mathbb{E}\left[\log\left(\eta_\infty(G_\theta)\right)\right]$$

$$\frac{1}{n}\sum_{p=1}^{n} \nabla_\theta \log \eta_n(G_\theta) \overset{n\to\infty}{\longrightarrow} \mathbb{E}\left[\nabla_\theta \log\left(\eta_\infty(G_\theta)\right)\right]$$

  where the expectation is taken over the invariant distribution of the Markov chain $\{X_n, \eta_n\}_{n\geq 0}$.
- Can use gradient descent as

$$\theta_{n+1} = \theta_n - \alpha_n \beta^{-1}\left[\nabla_\theta \log\left(\eta_n(G_\theta)\right)\right]_{\theta=\theta_n}.$$

SMC approximations for gradient based optimisation

- Will use a particle based method

$$\theta_{n+1} = \theta_n - \alpha_n \beta^{-1} \left(\widehat{\nabla_{\theta_n} \log}\,(Z_n) - \widehat{\nabla_{\theta_{n-1}} \log}\,(Z_{n-1}) \right)$$

- For the gradient can use Fisher identity

$$\begin{aligned}
\nabla_\theta \log Z_n &= \frac{1}{Z_n} \int \nabla_\theta \log \left(\prod_{p=1}^{n} G_\theta(x_p) M_\theta(x_{p-1}, dx_p) \nu(dx_0) \right) \\
&\quad \times G_\theta(x_p) M_\theta(x_{p-1}, dx_p) \nu(dx_0) \\
&= \frac{1}{Z_n} \int \left(\sum_{p=1}^{n} \frac{\nabla_\theta G_\theta(x_p)}{G_\theta(x_p)} + \frac{\nabla_\theta M_\theta(x_{p-1}, dx_p)}{M(x_{p-1}, dx_p)} \right) \\
&\quad \times G_\theta(x_p) M_\theta(x_{p-1}, dx_p) \nu(dx_0)
\end{aligned}$$

SMC Degeneracy

- Even using favourable assumptions like (A2) then the asymptotic variance of the standard SMC estimate $\widehat{I}_n^\theta$ of the additive functional

$$I_n^\theta = \frac{1}{Z_n} \int \left[\sum_{k=0}^{n} \varphi(x_k) \right] G_\theta(x_p) M_\theta(x_{p-1}, dx_p) \nu(dx_0), \qquad (3)$$

satisfies (Poyiadjis et al 2009 [5])

$$\mathbb{V}\left(\widehat{I}_n^\theta \right) \geq D_\theta \frac{n^2}{N}. \qquad (4)$$

- Even so one can obtain uniform L^p bounds for SMC approximations based on the marginals.

Smooth SMC Approximations

At time n, we start with the SMC approximation $\{\xi_n^i, \rho_n^i\}_{i=1}^L$ for the distribution flow μ_n

$$\widehat{\mu}_n(dx_n) = \sum_{i=1}^{L} \rho_n^i \delta_{\xi_n^i}(dx_n).$$

Then we can derive the following smooth approximations:

$$\widetilde{\eta}_{n+1}(dx) = \int \hat{\mu}_n(dx')M_\theta(x', dx)$$
$$= \sum_{i=1}^{L} \rho_n^i M_\theta(\xi_n^i, dx),$$

$$\widetilde{\mu}_n(dx) \propto G_\theta(x)\widetilde{\eta}_{n+1}(dx)$$
$$\propto \sum_{i=1}^{L} \rho_n^i G_\theta(x) M_\theta(\xi_n^i, dx).$$

Smooth SMC Approximations cont.

- We now propose new particles ξ^i_{n+1} from

$$\sum_{i=1}^{L} \rho^i_n Q_{n+1}(\xi^i_n, x)$$

  to obtain approximations $\widehat{\eta}_{n+1}$ and $\widehat{\mu}_{n+1}$.
- Q_{n+1} as in standard IS, i.e. chosen close to $M_\theta(x', x) G_\theta(x)$ and so that weights are well defined.
- Compute weights

$$\bar{\rho}^i_{n+1} = \frac{\widetilde{\eta}_{n+1}(\xi^i_{n+1})}{Q_{n+1}(\xi^i_n, \xi^i_{n+1})} = \frac{\sum_{i=1}^{L} \rho^i_n M_\theta(\xi^i_n, \xi^i_{n+1})}{Q_{n+1}(\xi^i_n, \xi^i_{n+1})},$$

$$w^i_{n+1} = \frac{\widetilde{\eta}_{n+1}(\xi^i_{n+1})}{Q_{n+1}(\xi^i_n, \xi^i_{n+1})} = \frac{\sum_{i=1}^{L} \rho^i_n G(\xi^i_{n+1}) M_\theta(\xi^i_n, \xi^i_{n+1})}{Q_{n+1}(\xi^i_n, \xi^i_{n+1})}, \ldots.$$

Smooth SMC Approximations cont.

$$\cdots \quad \rho^i_{n+1} = \frac{w^i_{n+1}}{\sum_{i=1}^{L} w^j_{n+1}}$$

- The corresponding SMC approximations

$$\widehat{\eta}_{n+1}(dx) = \sum_{i=1}^{L} \bar{\rho}^i_{n+1} \delta_{\xi^i_{n+1}}(dx),$$

$$\widehat{\mu}_{n+1}(dx) = \sum_{i=1}^{L} \rho^i_{n+1} \delta_{\xi^i_{n+1}}(dx).$$

$$\frac{\hat{Z}_{n+1}}{\hat{Z}_n} = \frac{1}{L} \sum_{i=1}^{L} w_n$$

Smooth SMC Approximations for the gradients

- Use marginal Fisher identity instead

$$\nabla_\theta \log Z_{n+1} = \int \nabla_\theta \log \left(G_\theta(x) \eta_{n+1}(x) \right) \mu_{n+1}(dx),$$

with smooth approximation $\widetilde{\nabla_\theta \log} \left(G_\theta(x) \eta_{n+1}(x) \right) =$

$$\frac{\sum_{i=1}^{L} \rho_n^i M_\theta(\xi_n^i, \xi_{n+1}^i) \left[\nabla_\theta \log G_\theta(x) + \nabla_\theta M_\theta(\xi_n^i, x) + \beta_n^i \right]}{\sum_{i=1}^{L} \rho_n^i M_\theta(\xi_n^i, \xi_{n+1}^i)}$$

where $\beta_n^i = \widetilde{\nabla_\theta \log} \left(G_{\theta_n}(\xi_n^i) \eta_n(\xi_n^i) \right).$

Smooth SMC Approximations for the gradients cont.

- At time n we start with the SMC approximation $\{\xi_n^i, \rho_n^i, \beta_n^i\}_{i=1}^L$ with $\beta_n^i = \widetilde{\nabla_\theta \log}\left(G_{\theta_n}(\xi_n^i)\eta_n(\xi_n^i)\right)$.
- Compute $\beta_{n+1}^i = \widetilde{\nabla_{\theta_n} \log}\left(G_\theta(\xi_{n+1}^i)\eta_{n+1}(\xi_{n+1}^i)\right)$.
- Compute $s_{n+1} = \sum_{i=1}^{L} \rho_{n+1}^i \beta_{n+1}^i$.
- Update parameter

$$\theta_{n+1} = \theta_n - \alpha_n \beta^{-1}\left(s_{n+1} - s_n\right).$$

Example revisited: Risk Sensitive LQR

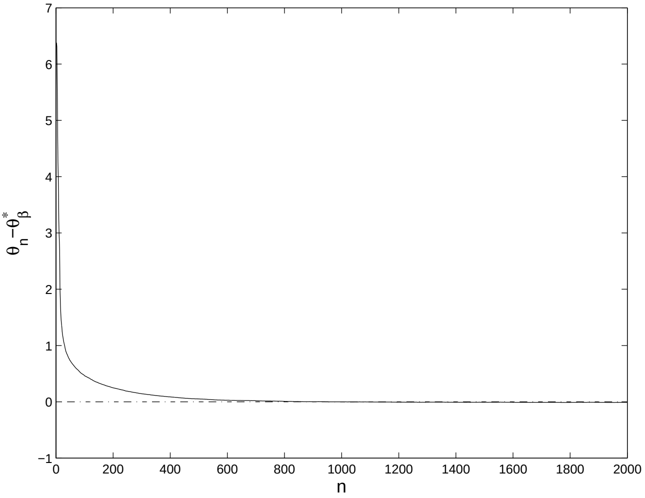


Figure: Plot of the error $\theta_n - \theta^*_\beta$ against n for $\beta = 0.001$, $\alpha_n = 0.01$, $L = 1000$, $\theta_0 = 5$.

Example revisited: Risk Sensitive LQR

	bias	bias	MSE	MSE
L	$\beta = 0.001$	$\beta = -0.001$	$\beta = 0.001$	$\beta = -0.001$
100	0.0263	0.0196	0.345	0.313
200	0.0141	0.0102	0.184	0.181
500	0.0067	0.0060	0.163	0.147
1000	0.0046	0.0039	0.123	0.109
2000	0.0036	0.0027	0.097	0.098
5000	0.0024	0.0018	0.081	0.080

Table: Observed absolute bias and total mean squared error (MSE) when computing estimates for θ^*_β for $\beta = \{0.001, -0.001\}$.

Conclusion

- Effectively we have used tools familiar online ML estimation but in a different context.
- Expensive ($\mathcal{O}(L^2)$ comp. cost), but added computation seems necessary to prevent degeneracy of standard SMC
- In preparation:
  - extension for the partially observed case
  - implementation for a more reallistic problem for portfolio optimisation

References

Del Moral P. (2004). *Feynman-Kac formulae: genealogical and interacting particle systems with applications*. New York: Springer Verlag.

Del Moral P., and Doucet A. (2004). Particle Motions in Absorbing Medium with Hard and Soft Obstacles. *Stochastic Analysis and Applications*. vol. 22, no. 5.

Di Masi G. B., Stettner L. (2000) Risk sensitive control of discrete time Markov processes with infinite horizon *SIAM J. Control Optimiz*, 38, 61 - 78.

Doucet, A., De Freitas, J.F.G. and Gordon N.J. (eds.) (2001). Sequential Monte Carlo Methods in Practice. New York: Springer-Verlag.

Poyadjis G., Doucet A. and Singh S.S., (2009) Sequential Monte Carlo for computing the score and observed information matrix in state-space models with applications to parameter estimation. Technical report CUED/F-INFENG/TR.628, Cambridge University.

Whittle, P. (1990) *Risk-Sensitive Optimal Control*. John Wiley and Sons Ltd.

Acknowledgements

Thanks for you attention!

The presenter is mostly grateful to the influence of