MATH-414 – Stochastic simulation

Lecture 6: Variance Reduction Techniques II

Prof. Fabio Nobile

Outline

Importance sampling

Importance sampling for stochastic processes

Importance sampling

- $\boldsymbol{X}$ random vector in $\mathbb{R}^d$ with (joint) pdf $f : \mathbb{R}^d \to \mathbb{R}_+$
- $Z = \psi(\boldsymbol{X}) \in \mathbb{R}$ output quantity
- **Goal**: compute $\mu = \mathbb{E}[Z] = \mathbb{E}[\psi(\boldsymbol{X})]$

Consider an auxiliary pdf g s.t. $g(\boldsymbol{x}) = 0 \Rightarrow \psi(\boldsymbol{x})f(\boldsymbol{x}) = 0$. Then

$$\mu = \mathbb{E}[Z] = \int_{\mathbb{R}^d} \left(\frac{\psi(\boldsymbol{x})f(\boldsymbol{x})}{g(\boldsymbol{x})} \right) g(\boldsymbol{x})\, d\boldsymbol{x} = \mathbb{E}\left[\frac{\psi(\tilde{\boldsymbol{X}})f(\tilde{\boldsymbol{X}})}{g(\tilde{\boldsymbol{X}})} \right], \quad \tilde{\boldsymbol{X}} \sim g$$

Importance sampling Monte Carlo estimator:

- generate N iid replicas $\tilde{\boldsymbol{X}}^{(i)} \sim g$
- compute $\hat{\mu}_{IS} = \frac{1}{N} \sum_{i=1}^{N} \frac{\psi(\tilde{\boldsymbol{X}}^{(i)})f(\tilde{\boldsymbol{X}}^{(i)})}{g(\tilde{\boldsymbol{X}}^{(i)})}$

Nomenclature

- g: importance sampling distribution (or dominating distribution)
- $w(\boldsymbol{x}) = \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}$: likelihood ratio

Importance sampling – algorithm

Algorithm: Importance sampling

1. Generate N iid replicas $\tilde{X}^{(1)}, \ldots, \tilde{X}^{(N)} \sim g$
2. Compute $\hat{\mu}_{\mathsf{IS}} = \frac{1}{N} \sum_{i=1}^{N} \psi(\tilde{X}^{(i)}) w(\tilde{X}^{(i)})$, $\qquad w(\tilde{X}^{(i)}) = \frac{f(\tilde{X}^{(i)})}{g(\tilde{X}^{(i)})}$
3. Estimate $\hat{\sigma}^2_{\mathsf{IS}} = \frac{1}{N-1} \sum_{i=1}^{N} \left(\psi(\tilde{X}^{(i)}) w(\tilde{X}^{(i)}) - \hat{\mu}_{\mathsf{IS}} \right)^2$
4. Output $\hat{\mu}_{\mathsf{IS}}$ and a (asymptotic) $1-\alpha$ confidence interval

$$\hat{I}_{\alpha,N} = \left[\hat{\mu}_{\mathsf{IS}} - c_{1-\alpha/2} \frac{\hat{\sigma}_{\mathsf{IS}}}{\sqrt{N}}, \;\; \hat{\mu}_{\mathsf{IS}} + c_{1-\alpha/2} \frac{\hat{\sigma}_{\mathsf{IS}}}{\sqrt{N}} \right]$$

Example – option pricing

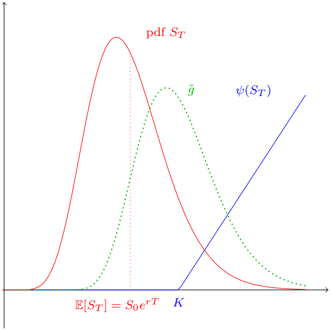
- S_t: value of an asset at time t, modeled by

$$dS_t = rS_t\,dt + \sigma S_t\,dW_t, \quad t \in (0, T]$$

- call option: payoff $\psi(S_T) = (S_T - K)_+$:
- **Goal**: estimate (discounted) option price $\mu = \mathbb{E}\left[e^{-rT}\psi(S_T)\right]$

- If $K \gg S_0$ only few samples will fall above the strike price $\rightsquigarrow$ CMC estimator is inefficient
- **Idea**: increase interest rate r to enhance the probability of S_T being above the strike

$$d\tilde{S}_t = \tilde{r}\tilde{S}_t\,dt + \sigma\tilde{S}_t\,dW_t$$



Example – option pricing

- original random variable: $S_T = S_0 e^{X_T}$, with $X_T \sim N((r - \sigma^2/2)T, \sigma^2 T)$
- option price: $\mu = \mathbb{E}\left[\tilde{\psi}(X_T)\right]$, with $\tilde{\psi}(X_T) = e^{-rT}(S_0 e^{X_T} - K)_+$
- modified random variable: $\tilde{S}_T = S_0 e^{\tilde{X}_T}$, with $\tilde{X}_T \sim N((\tilde{r} - \sigma^2/2)T, \sigma^2 T)$
- likelihood ratio: $w(x) = \frac{f_{X_T}(x)}{f_{\tilde{X}_T}(x)} = \exp\left\{\frac{(\tilde{r}-r)((\tilde{r}+r-\sigma^2)T-2x)}{2\sigma^2}\right\}$
- Importance sampling

$$\mu = \mathbb{E}\left[\tilde{\psi}(\tilde{X}_T)w(\tilde{X}_T)\right]$$

with $\tilde{X}_T$ following the modified GBM.

Choice of importance sampling distribution

Consider the importance sampling Monte Carlo estimator

$$\hat{\mu}_{\mathsf{IS}} = \frac{1}{N}\sum_{i=1}^{N} \frac{\psi(\tilde{X}^{(i)}) f(\tilde{X}^{(i)})}{g(\tilde{X}^{(i)})}, \qquad \tilde{X}^{(i)} \overset{\text{iid}}{\sim} g$$

- $\hat{\mu}_{IS}$ is unbiased.
- Variance of $\hat{\mu}_{IS}$:

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[\hat{\mu}_{\mathsf{IS}}\right] &= \frac{1}{N}\mathbb{V}\mathrm{ar}_g\left(\frac{\psi f}{g}\right) = \frac{1}{N}\left(\int_{\mathbb{R}^d} \frac{\psi^2(x) f^2(x)}{g^2(x)} g(x)\, dx - \mu^2\right) \\ &= \frac{1}{N}\left(\mathbb{E}_f\left[\psi^2 \frac{f}{g}\right] - \mu^2\right)\end{aligned}$$

Can we choose optimally g to minimize the variance of the extimator?

Choice of importance sampling distribution

Constrained minimization problem

$$\min_g \int \frac{\psi^2(x) f^2(x)}{g(x)} dx \quad \text{s.t.} \quad \int g(x) dx = 1, \quad g \geq 0$$

Lagrangian multiplier approach: $\mathcal{L}(g, \lambda) = \int_\Gamma \frac{\psi^2 f^2}{g} \, dx + \lambda \left(\int_\Gamma g - 1 \right)$
taking variations:

$$\frac{\partial \mathcal{L}}{\partial g}(\delta g) = -\int_{\mathbb{R}^d} \left(\psi^2 \frac{f^2}{g^2} - \lambda \right) \delta g dx = 0, \qquad \forall \delta g$$

$$\Longrightarrow \qquad g^2 = \frac{\psi^2 f^2}{\lambda}$$

optimal pdf: $g^* = \frac{|\psi| f}{\int |\psi| f}$

- Not practical: Normalizing constant $\int |\psi| f dx$ not know (and as difficult to compute as $\mathbb{E}_{X \sim f}[\psi(X)]$)
- Gives guidelines on how to construct good importance sampling distributions

Optimal distribution over a parametric family

variance minimization

- Let $\mathcal{F} = \{f(\cdot, \theta),\ \theta \in \Theta\}$ be a paramtric family of distributions (e.g. exponenatial family)
- Assume that the initial distribution f is in $\mathcal{F}$, i.e. $f(\cdot) = f(\cdot, \theta_0)$

Idea: look for optimal g within $\mathcal{F}$:

$$g(\cdot) = f(\cdot, \theta^*), \qquad \text{with } \theta^* = \operatorname*{argmin}_{\theta \in \Theta} \mathbb{E}_{\theta_0}\left[\frac{\psi^2 f(\cdot, \theta_0)}{f(\cdot, \theta)}\right].$$

Algorithm: Importance sampling with variance minimization

---

1. Generate $\bar{N}$ iid replicas $Y^{(1)}, \dots, Y^{(\bar{N})} \sim f(\cdot, \theta_0)$
2. Solve the minimization problem

$$\hat{\theta}^*_{\mathbf{Y}} = \operatorname*{argmin}_{\theta \in \Theta} \frac{1}{\bar{N}} \sum_{i=1}^{\bar{N}} \psi^2(Y^{(i)}) \frac{f(Y^{(i)}, \theta_0)}{f(Y^{(i)}, \theta)}$$

3. Generate N iid replicas $X^{(1)}, \dots, X^{(N)} \sim f(\cdot, \hat{\theta}^*_{\mathbf{Y}})$
4. Compute $\hat{\mu}_{\text{IS}} = \frac{1}{N} \sum_{i=1}^{N} \psi(X^{(i)}) \frac{f(X^{(i)}, \theta_0)}{f(X^{(i)}, \hat{\theta}^*_{\mathbf{Y}})}$.

Adaptive importance sampling

The previous algorithm can be made adaptive:

- suppose that at the $(k-1)$-th iteration we have estimated the parameter $\theta^{(k-1)}$
- Then, at iteration k we generate from $f(\cdot, \theta^{(k-1)})$ and we have to minimize the variance

$$\theta^{(k)} = \underset{\theta}{\operatorname{argmin}}\, \mathbb{E}_{\theta_0}\left[\frac{\psi^2 f(\cdot,\theta_0)}{f(\cdot,\theta)}\right] = \mathbb{E}_{\theta^{(k-1)}}\left[\frac{\psi^2 f^2(\cdot,\theta_0)}{f(\cdot,\theta) f(\cdot,\theta^{(k-1)})}\right]$$

Algorithm: Adaptive importance sampling with variance minimization

Given: *tol*, α, θ_0, $\bar{N} > 1$, $\gamma > 1$

Set $N = \bar{N}$, $\hat{\theta} = \theta_0$, $\hat{\sigma} = \infty$

while $\frac{\hat{\sigma} c_{1-\alpha/2}}{\sqrt{N}} > tol$ **do**

  Generate N iid replicas $Y^{(1)}, \ldots, Y^{(N)} \sim f(\cdot, \hat{\theta})$

  Compute $\hat{\mu}_{IS} = \frac{1}{N}\sum_{i=1}^N \psi(Y^{(i)}) \frac{f(Y^{(i)},\theta_0)}{f(Y^{(i)},\hat{\theta})}$

  Optimize $\hat{\theta}_{new} = \operatorname{argmin}_{\theta\in\Theta} \frac{1}{N}\sum_{i=1}^N \psi^2(Y^{(i)}) \frac{f^2(Y^{(i)},\theta_0)}{f(Y^{(i)},\theta) f(Y^{(i)},\hat{\theta})}$

  Set $\hat{\theta} = \hat{\theta}_{new}$ and $N = \gamma N$

end

Output $\hat{\mu}_{\text{IS}}$

Optimal distribution over a parametric family

cross entropy minimization

Definition. *The Kullback-Leibler divergence $D_{KL}(g|f)$ between a target pdf g and a candidate pdf f is defined as*

$$D_{KL}(g|f) = \mathbb{E}_g[\log \frac{g}{f}] = \int g(x) \log g(x) dx - \int g(x) \log f(x) dx.$$

$D_{KL}(g|f) \geq 0$ and $D_{KL}(g|f) = 0$ if and only if $g = f$ a.e.

Idea: find the pdf $f \in \mathcal{F}$ that minimizes the KL divergence to the optimal importance sampling distribution $g^* = \frac{|\psi| f}{\int \| \psi | f}$

$$\begin{aligned}
\theta^* &= \operatorname*{argmin}_\theta D_{KL}(g^* | f(\cdot, \theta)) = \operatorname*{argmin}_\theta \mathbb{E}_{g^*}[\log g^*] - \mathbb{E}_{g^*}[\log f(\cdot, \theta)] \\
&= \operatorname*{argmax}_\theta \frac{1}{\mathbb{E}_{\theta_0}[|\psi|]} \int |\psi(x)| f(x, \theta_0) \log f(x, \theta) dx \\
&= \operatorname*{argmax}_\theta \int |\psi(x)| f(x, \theta_0) \log f(x, \theta) dx \\
&= \operatorname*{argmax}_\theta \mathbb{E}_{\hat{\theta}} \left[|\psi(\cdot)| \frac{f(\cdot, \theta_0)}{f(\cdot, \hat{\theta})} \log f(\cdot, \theta) \right]
\end{aligned}$$

Adaptive cross-entropy importance sampling

Algorithm: Adaptive importance sampling with cross entropy minimization

Given: tol, α, θ_0, $\bar{N} > 1$, $\gamma > 1$

Set $N = \bar{N}$, $\hat{\theta} = \theta_0$, $\hat{\sigma} = \infty$

while $\frac{\hat{\sigma} c_{1-\alpha/2}}{\sqrt{N}} > tol$ **do**

  Generate N iid replicas $Y^{(1)}, \ldots, Y^{(N)} \sim f(\cdot, \hat{\theta})$

  Compute $\hat{\mu}_{IS} = \frac{1}{N} \sum_{i=1}^{N} \psi(Y^{(i)}) \frac{f(Y^{(i)}, \theta_0)}{f(Y^{(i)}, \hat{\theta})}$

  Optimize $\hat{\theta}_{new} = \text{argmax}_{\theta} \frac{1}{N} |\psi(Y^{(i)})| \frac{f(Y^{(i)}, \theta_0)}{f(Y^{(i)}, \hat{\theta})} \log f(Y^{(i)}, \theta)$

  Set $\hat{\theta} = \hat{\theta}_{new}$ and $N = \gamma N$

end

Output $\hat{\mu}_{\text{IS}}$

Weighted importance sampling

In certain cases, the pdf f and/or the dominating pdf g, are known only up to a normalizing constant.

Let $f = c_g \tilde{f}$ and $g = c_g \tilde{g}$, with $c_f = (\int \tilde{f})^{-1}$ and $c_g = (\int \tilde{g})^{-1}$.

Self-normalized importance sampling estimator

$$\hat{\mu}_{\mathrm{IS}}^{W} = \frac{\sum_{i=1}^{N} \psi(X^{(i)}) w(X^{(i)})}{\sum_{i=1}^{N} w(X^{(i)})}, \quad \text{with} \quad w(x) = \frac{\tilde{f}(x)}{\tilde{g}(x)}, \quad X^{(i)} \overset{\text{iid}}{\sim} g$$

The estimator $\hat{\mu}_{\mathrm{IS}}^{W}$ is asymptotically consistent. Indeed by SLLN

$$\frac{1}{N} \sum_{i=1}^{N} w(X^{(i)}) \xrightarrow{\text{a.s.}} \int \frac{\tilde{f}(x)}{\tilde{g}(x)} g(x)\, dx = \frac{c_g}{c_f}$$

$$\frac{1}{N} \sum_{i=1}^{N} \psi(X^{(i)}) w(X^{(i)}) \xrightarrow{\text{a.s.}} \int \psi \frac{\tilde{f}}{\tilde{g}} g\, dx = \frac{c_g}{c_f} \mu$$

It is however, biased, in general, although the bias is usually small.

Importance sampling for stochastic processes

Discrete time Markov Chains

Consider a discrete time Markov chain $\{X_n,\ n \in \mathbb{N}_0\} \sim \mathrm{Markov}\,(p_0, P)$ in $\mathbb{R}^d$, with transition density function $p : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$:, i.e.

$$P(x, A) = \mathbb{P}\,(X_{n+1} \in A \mid X_n = x) = \int_A p(x, y)\, dy, \quad A \in \mathcal{B}(\mathbb{R}^d),$$

Goal: compute $\mu = \mathbb{E}\,[Z] = \mathbb{E}\,[\psi(X_{0:m})]$

where $X_{0:m} = (X_0, \ldots, X_m)$ corresponds to the path up to step m.

Question: how to do importance sampling in this case?

Take dominating densities

$$\begin{aligned} &q_0 \gg p_0 && \text{(i.e. } q_0(y) = 0 \implies p_0(y) = 0,\ \forall y) \\ &q(x, \cdot) \gg p(x, \cdot),\ \forall x && \text{(i.e. } q(x, y) = 0 \implies p(x, y) = 0,\ \forall y) \end{aligned}$$

Shorthand notation:

$$\begin{aligned} \{X_n\} &\sim p_0, P && \text{if } \{X_n\} \sim \mathrm{Markov}\{p_0, P\} \\ \{X_n\} &\sim q_0, Q && \text{if } \{X_n\} \sim \mathrm{Markov}\{q_0, Q\} \end{aligned}$$

Importance sampling for stochastic processes

Discrete time Markov Chains

$$\mu = \mathbb{E}_{X_{0:m}\sim p_0,P}[\psi(X_{0:m})] = \int \psi(x_0,\dots,x_m)p_{X_{0:m}}(x_0,\dots,x_m)dx_0\cdots dx_m$$

$$= \int \psi(x_0,\dots,x_m)p_0(x_0)p(x_0,x_1)\cdots p(x_{m-1},x_m)dx_0\cdots dx_m$$

$$= \int \psi(x_0,\dots,x_m)\frac{p_0(x_0)\prod_{i=1}^m p(x_{i-1},x_i)}{q_0(x_0)\prod_{i=1}^m q(x_{i-1},x_i)}q_0(x_0)\prod_{i=1}^m q(x_{i-1},x_i)dx_0\cdots dx_m$$

$$= \mathbb{E}_{X_{0:m}\sim q_0,Q}[\psi(X_{0:m})w(X_{0:m})]$$

with likelihood ratio $w(X_{0:m}) = \frac{p_0(X_0)}{q_0(X_0)}\prod_{i=1}^m \frac{p(X_{i-1},X_i)}{q(X_{i-1},X_i)}$.

The previous formula generalizes to the case of a stopped process. Let τ be a stopping time (e.g. $\tau = \min\{n\geq 0: X_n\in A\}$) and aim to compute $\mu = \mathbb{E}[\psi_\tau(X_{0:\tau})\mathbb{1}_{\tau<\infty}]$ with $\{X_n\}\sim \mathrm{Markov}(p_0,P)$. Then

$$\mu = \mathbb{E}_{\{X_n\}\sim q_0,Q}[\psi_\tau(X_{0:\tau})\mathbb{1}_{\tau<\infty}w(X_{0:\tau})] \quad \text{with } w(X_{0:\tau}) \text{ as above}$$

as long as $\tau<\infty$ under P $\implies$ $\tau<\infty$ under Q.

Importance sampling for stochastic processes

Discrete time Markov Chains

Algorithm: Importance sampling for Markov processes.

1 Generate N iid paths $X^{(i)}_{0:\tau^{(i)}} = (X^{(i)}_0, \dots, X^{(i)}_{\tau^{(i)}})$, $i = 1, \dots, N$, each one up to the stopping time $\tau^{(i)}$, of the Markov chain with transition probability $q : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ and initial probability $q_0 : \mathbb{R}^d \to \mathbb{R}_+$

2 Compute likelihood ratio $w(X^{(i)}_{0:\tau^{(i)}}) = \dfrac{p_0(X^{(i)}_0)}{q_0(X^{(i)}_0)} \displaystyle\prod_{k=1}^{\tau^{(i)}} \dfrac{p(X^{(i)}_{k-1}, X^{(i)}_k)}{q(X^{(i)}_{k-1}, X^{(i)}_k)}$

3 Compute $\hat{\mu}_{\mathsf{IS}} = \frac{1}{N}\sum_{i=1}^{N} \psi_{\tau^{(i)}}(X^{(i)}_{0:\tau^{(i)}}) w(X^{(i)}_{0:\tau^{(i)}})$

4 Output $\hat{\mu}_{\mathsf{IS}}$ and a confidence interval based on $\hat{\sigma}_{\mathsf{IS}}$.

Importance sampling for stochastic processes

Discretized stochastic differential equations

Consider a stochastic differential equation in $\mathbb{R}^d$

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dW_t, \quad t > 0, \qquad \text{with } X_0 \text{ given}, \quad (1)$$

- W_t: d-dimensional Brownian motion
- $b : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$: drift
- $\sigma : \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^{d\times d}$: diffusion matrix

Goal: compute $\mu = \mathbb{E}[Z] = \mathbb{E}[\psi(\{X_t\}_{0\le t\le T})]$

(e.g. $Z = \int_0^T X_{s,1} ds$, $\quad Z = \|X_T\|$, $\quad$ etc.)

Discretization by Euler Maruyama:

$$X_{n+1} = X_n + b(X_n, t_n)\Delta t + \sigma(X_n, t_n)\xi_n, \quad \xi_n \sim N(0, I_{d\times d}\Delta t).$$

Discretized output

$$\mu_{\Delta t} = \mathbb{E}[\psi_{\Delta t}(X_0, \ldots, X_m)] = \mathbb{E}_{\xi_0,\ldots,\xi_{m-1}}[\hat{\psi}(\xi_0, \ldots, \xi_{m-1})]$$

Importance sampling for stochastic processes

Discretized stochastic differential equations

How to do importance sampling in this case? Idea: change the drift of the SDE to $\tilde{b}(X_n, t_n)$.

This corresponds to changing the mean of the Gaussian increments:

$$\tilde{\xi}_n \sim N(\phi(X_n, t_n)\Delta t, I_{d\times d}\Delta t), \quad \phi(X_n, t_n) = \sigma^{-1}(X_n, t_n)(\tilde{b}(X_n, t_n) - b(X_n, t_n))$$

Indeed, writing $\tilde{\xi}_n = \phi(X_n, t_n)\Delta t + \eta_n$, with $\eta_n \sim N(0, I_{d\times d}\Delta t)$ we have

$$\begin{aligned} X_{n+1} &= X_n + b(X_n, t_n)\Delta t + \sigma(X_n, t_n)\tilde{\xi}_n \\ &= X_n + \tilde{b}(X_n, t_n)\Delta t + \sigma(X_n, t_n)\eta_n \end{aligned}$$

We then have

$$\mu_{\Delta t} = \mathbb{E}_{\xi_{0:m-1}}[\hat{\psi}(\xi_{0:m-1})] = \mathbb{E}_{\tilde{\xi}_{0:m-1}}\left[\hat{\psi}(\tilde{\xi}_{0:m-1}) w(\tilde{\xi}_{0:m-1})\right]$$

Denoting $z \mapsto p(z; \mu, \Sigma)$ the joint pdf of a Gaussian vector with mean μ and covariance matrix Σ, the likelihood ratio reads

$$
\begin{aligned}
w(\tilde{\xi}_{0:m-1}) &= \prod_{i=0}^{m-1} \frac{p(\tilde{\xi}_i; 0, I_{d\times d}\Delta t)}{p(\tilde{\xi}_i; \phi(X_i, t_i)\Delta t, I_{d\times d}\Delta t)} \\
&= \prod_{i=0}^{m-1} \exp\left(-\frac{1}{2\Delta t}\|\tilde{\xi}_i\|^2 + \frac{1}{2\Delta t}\|\tilde{\xi}_i - \phi(X_i, t_i)\Delta t\|^2 \right) \\
&= \prod_{i=0}^{m-1} \exp\left(\frac{\Delta t}{2}\|\phi(X_i, t_i)\|^2 - \phi(X_i, t_i)^T \tilde{\xi}_i \right) \\
&= \exp\left(\frac{1}{2}\sum_{i=0}^{m-1} \Delta t \|\phi(X_i, t_i)\|^2 - \sum_{i=0}^{m-1} \phi(X_i, t_i)^T \tilde{\xi}_i \right)
\end{aligned}
$$

Importance sampling for stochastic processes

Discretized stochastic differential equations

Algorithm: Importance sampling for SDEs.

1 Generate N iid paths $X_{0:m}^{(i)}$, $i = 1, \ldots, N$ with modified drift

$$X_{n+1}^{(i)} = X_n^{(i)} + b(X_n^{(i)}, t_n)\Delta t + \sigma(X_n^{(i)}, t_n)\tilde{\xi}_n^{(i)}, \quad \tilde{\xi}_n^{(i)} \sim N(\phi(X_n, t_n)\Delta t, I_{d\times d}\Delta t) \tag{2}$$

2 Compute likelihood ratio

$$w(\tilde{\xi}_{0:m-1}^{(i)}) = \exp\left(\frac{1}{2}\sum_{n=0}^{m-1}\Delta t\|\phi(X_n^{(i)}, t_n)\|^2 - \sum_{n=0}^{m-1}\phi(X_n, t_n)^T\tilde{\xi}_n^{(i)}\right)$$

3 Compute $\hat{\mu}_{\mathrm{IS}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\psi}(\tilde{\xi}_{0:m-1}^{(i)})w(\tilde{\xi}_{0:m-1}^{(i)})$

4 Output $\hat{\mu}_{\mathrm{IS}}$ and a confidence interval based on $\hat{\sigma}_{\mathrm{IS}}$.

Importance sampling for stochastic processes

Discretized stochastic differential equations

In the limit $\Delta t \to 0$ we can define a drifted Brownian motion $d\tilde{W}_t = \phi(X_i, t_i)dt + dW_t$ (with W_t a standad BM) and

$$w(\{\tilde{W}_t\}_{0\leq t\leq T}) = \exp\left(\frac{1}{2}\int_0^T \|\phi(X_t,t)\|^2 dt - \int_0^T \phi(X_t,t)\cdot d\tilde{W}_t\right)$$

This represents the ratio between the (joint) densities of W_t and $\tilde{W}_t$ denoted $\frac{d\mathbb{P}_{W_t}}{d\mathbb{P}_{\tilde{W}_t}}$ (Girsanov's theorem)

Then we can write (at least formally)

$$\mu = \mathbb{E}_{W_t}[\psi(\{X_t\}_{0\leq t\leq T})] = \mathbb{E}_{\tilde{W}_t}\left[\psi(\{X_t\}_{0\leq t\leq T})\frac{d\mathbb{P}_{W_t}}{d\mathbb{P}_{\tilde{W}_t}}(\tilde{W}_t)\right].$$

Importance sampling for stochastic processes

Continuous time discrete space Markov processes

Consider a continuous time Markov process taking values in the discrete space $\mathcal{X} = \{y_1, y_2, \dots\}$

$$\{X_t \in \mathcal{X},\ t \geq 0\} \sim \mathrm{Markov}(\lambda, Q)$$

(Q - stable and conservative generator matrix; λ initial distribution)

Goal: compute $\mu = \mathbb{E}[Z] = \mathbb{E}[\psi(\{X_t\}_{0\leq t\leq T})]$

Importance sampling by changing (λ, Q) into $(\tilde{\lambda}, \tilde{Q})$. Then

$$\mu = \mathbb{E}_{\lambda,Q}[\psi(\{X_t\}_{0\leq t\leq T})] = \mathbb{E}_{\tilde{\lambda},\tilde{Q}}[\psi(\{X_t\}_{0\leq t\leq T}) w(\{X_t\}_{0\leq t\leq T})]$$

Denoting $\{J_n\}$ the jump times; $\{S_n\}$ the holding times; $\{Y_n\}$ the visited states, then the Likelihood ratio reads

$$w(\{X_t\}_{0\leq t\leq T}) = \frac{\lambda_{X_0}}{\tilde{\lambda}_{X_0}} \left(\prod_{i=1}^{N(T)} \frac{Q_{Y_{i-1}Y_i}}{\tilde{Q}_{Y_{i-1}Y_i}} \frac{\exp\{-S_j Q_{Y_{j-1}}\}}{\exp\{-S_j \tilde{Q}_{Y_{j-1}}\}} \right) \frac{\exp\{-(T - J_{N(T)}) Q_{Y_{N(T)}}\}}{\exp\{-(T - J_{N(T)}) \tilde{Q}_{Y_{N(T)}}\}}$$

$$= \frac{\lambda_{X_0}}{\tilde{\lambda}_{X_0}} \left(\prod_{i=1}^{N(T)} \frac{Q_{Y_{i-1}Y_i}}{\tilde{Q}_{Y_{i-1}Y_i}} \right) \exp\left\{ -\int_0^T (Q_{Y_s} - \tilde{Q}_{Y_s}) ds \right\}$$

with $Q_i = -Q_{ii} = \sum_{j\neq i} Q_{ij}$.