Weierstrass Institute for
Applied Analysis and Stochastics

Leibniz Association

Simulation of conditional diffusions via forward-reverse stochastic representations

Christian Bayer and John Schoenmakers

SRI UQ Workshop 2015, KAUST

Mohrenstrasse 39 · 10117 Berlin · Germany · Tel. +49 30 20372 0 · www.wias-berlin.de
January 6, 2015

Outline

Given $a : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$, a standard, m-dimensional Brownian motion B, consider

$$dX(s) = a(s, X(s))ds + \sigma(s, X(s))dB(s), \quad 0 \le s \le T$$

Goal

Given a grid $\mathcal{D} = \{0 = s_0 < \cdots < s_{K+L+1} = T\}$, $f : \mathbb{R}^{(K+L)d} \to \mathbb{R}$ and $x, y \in \mathbb{R}^d$, compute

$$\mathbb{E}\left[f\left(X(s_1), \ldots, X(s_{K+L})\right) \mid X(0) = x,\ X(T) = y \right].$$

Given $a : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$, a standard, m-dimensional Brownian motion B, consider

$$dX(s) = a(s, X(s))ds + \sigma(s, X(s))dB(s), \quad 0 \le s \le T$$

Goal (extended)

Given a grid $\mathcal{D} = \{ 0 = s_0 < \cdots < s_{K+L+1} = T \}$, $f : \mathbb{R}^{(K+L)d} \to \mathbb{R}$ and $A \subset \mathbb{R}^d$, compute

$$\mathbb{E}\left[f(X(s_1), \ldots, X(s_{K+L})) \mid X(0) = x,\ X(T) \in A \right],$$

A with positive measure or d'-dimensional hyperplanes, $0 \le d' \le d$.

Given $a : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$, a standard, m-dimensional Brownian motion B, consider

$$dX(s) = a(s, X(s))ds + \sigma(s, X(s))dB(s), \quad 0 \le s \le T$$

Goal

Given a grid $\mathcal{D} = \{0 = s_0 < \cdots < s_{K+L+1} = T\}$, $f : \mathbb{R}^{(K+L)d} \to \mathbb{R}$ and $x, y \in \mathbb{R}^d$, compute

$$\mathbb{E}\left[f(X(s_1), \ldots, X(s_{K+L})) \mid X(0) = x,\ X(T) = y\right].$$

Extension

- Discrete time Markov chains [B., Mai, Schoenmakers]
- Pure jump Markov process [B., Moraes, Villanova, Tempone]

- Algorithm for maximizing likelihood with missing data

Example: Two-stage hierarchical model:

- Random variables Y and U (multi-variate)
- $U \sim h(\cdot;\theta)$, $Y|U = u \sim f(\cdot|u;\theta)$, $\theta \in \Theta$
- Data: y (instance of Y), but U not observable

Algorithm

Let $l(\theta;y) := \log \int f(y|u;\theta)h(u;\theta)du$, $\hat{\theta} := \arg\max_{\theta\in\Theta} l(\theta;y)$. *Given* θ_0.

(E) $Q(\theta|\theta_n, y) := E_{\theta_n}\left[\log\left(f(y|U,\theta)h(U;\theta)\right) \mid Y = y\right]$

(M) $\theta_{n+1} := \arg\max_{\theta\in\Theta} Q(\theta|\theta_n, y)$.

- $l(\theta_{n+1};y) \geq l(\theta_n;y)$
- Weak conditions: $\theta_n \to \theta^*$ with $\nabla l(\theta^*;y) = 0$

- Algorithm for maximizing likelihood with missing data

Example: Two-stage hierarchical model:

- Random variables Y and U (multi-variate)
- $U \sim h(\cdot;\theta)$, $Y|U = u \sim f(\cdot|u;\theta)$, $\theta \in \Theta$
- Data: y (instance of Y), but U not observable

Algorithm

Let $l(\theta;y) := \log \int f(y|u;\theta)h(u;\theta)du$, $\hat{\theta} := \arg\max_{\theta\in\Theta} l(\theta;y)$. *Given* θ_0.

(E) $Q(\theta|\theta_n, y) := E_{\theta_n}\left[\log\left(f(y|U,\theta)h(U;\theta)\right) \mid Y = y\right]$

(M) $\theta_{n+1} := \arg\max_{\theta\in\Theta} Q(\theta|\theta_n, y)$.

- $l(\theta_{n+1};y) \geq l(\theta_n;y)$
- Weak conditions: $\theta_n \to \theta^*$ with $\nabla l(\theta^*;y) = 0$

- Algorithm for maximizing likelihood with missing data

Example: Two-stage hierarchical model:

- Random variables Y and U (multi-variate)
- $U \sim h(\cdot;\theta)$, $Y|U = u \sim f(\cdot|u;\theta)$, $\theta \in \Theta$
- Data: y (instance of Y), but U not observable

Algorithm

Let $l(\theta; y) := \log \int f(y|u;\theta)h(u;\theta)du$, $\hat{\theta} := \arg\max_{\theta\in\Theta} l(\theta;y)$. *Given* θ_0.

(E) $Q(\theta|\theta_n, y) := E_{\theta_n}\left[\log\left(f(y|U,\theta)h(U;\theta)\right) \mid Y = y\right]$

(M) $\theta_{n+1} := \arg\max_{\theta\in\Theta} Q(\theta|\theta_n, y)$.

- $l(\theta_{n+1}; y) \geq l(\theta_n; y)$
- Weak conditions: $\theta_n \longrightarrow \theta^*$ with $\nabla l(\theta^*; y) = 0$

- Algorithm for maximizing likelihood with missing data

Example: Two-stage hierarchical model:

- Random variables Y and U (multi-variate)
- $U \sim h(\cdot;\theta)$, $Y|U = u \sim f(\cdot|u;\theta)$, $\theta \in \Theta$
- Data: y (instance of Y), but U not observable

Algorithm

Let $l(\theta; y) := \log \int f(y|u;\theta)h(u;\theta)du$, $\hat{\theta} := \arg\max_{\theta\in\Theta} l(\theta; y)$. *Given* θ_0.

(E) $Q(\theta|\theta_n, y) := E_{\theta_n}\left[\log\left(f(y|U,\theta)h(U;\theta)\right) \,|\, Y = y\right]$

(M) $\theta_{n+1} := \arg\max_{\theta\in\Theta} Q(\theta|\theta_n, y)$.

- $l(\theta_{n+1}; y) \geq l(\theta_n; y)$
- Weak conditions: $\theta_n \to \theta^*$ with $\nabla l(\theta^*; y) = 0$

- OU-process: $dX_s = -\theta X_s ds + dW_s, \quad s \in [0, T]$
- On path space:

$$L_c(X;\theta) = \frac{dP^\theta}{dP^0}(X) = \exp\left(-\theta \int_0^T X_s dX_s - \frac{\theta^2}{2} \int_0^T X_s^2 ds\right)$$

- Discrete observations: $\mathbf{x} = (x_0, \ldots, x_K)$ of $\mathbf{X} := (X(s_0), \ldots, X(s_K))$, $s_0 = 0,\ s_K = T$
- EM algorithm with

$$Q(\theta|\theta_n, \mathbf{x}) = \mathbb{E}_{\theta_n}\left[-\theta \int_0^T X_s dX_s - \frac{\theta^2}{2} \int_0^T X_s^2 ds \,\middle|\, \mathbf{X} = \mathbf{x}\right]$$

$$= \sum_{i=1}^K \mathbb{E}_{\theta_n}\left[-\theta \int_{s_{i-1}}^{s_i} X_s dX_s - \frac{\theta^2}{2} \int_{s_{i-1}}^{s_i} X_s^2 ds \,\middle|\, X_{s_{i-1}} = x_{i-1}, X_{s_i} = x_i\right]$$

- OU-process: $dX_s = -\theta X_s ds + dW_s, \quad s \in [0,T]$
- On path space:

$$L_c(X;\theta) = \frac{dP^\theta}{dP^0}(X) = \exp\left(-\theta \int_0^T X_s dX_s - \frac{\theta^2}{2}\int_0^T X_s^2 ds\right)$$

- Discrete observations: $\mathbf{x} = (x_0, \dots, x_K)$ of $\mathbf{X} := (X(s_0), \dots, X(s_K))$, $s_0 = 0,\ s_K = T$
- EM algorithm with

$$Q(\theta|\theta_n, \mathbf{x}) = \mathbb{E}_{\theta_n}\left[-\theta \int_0^T X_s dX_s - \frac{\theta^2}{2}\int_0^T X_s^2 ds \;\middle|\; \mathbf{X} = \mathbf{x}\right]$$

$$= \sum_{i=1}^K \mathbb{E}_{\theta_n}\left[-\theta \int_{s_{i-1}}^{s_i} X_s dX_s - \frac{\theta^2}{2}\int_{s_{i-1}}^{s_i} X_s^2 ds \;\middle|\; X_{s_{i-1}} = x_{i-1}, X_{s_i} = x_i\right]$$

- Notation: $X_{t,x}(s)$ solution of SDE started at $X_{t,x}(t) = x$, $t < s$
- Generator of the SDE:

$$L_t f(x) = \langle \nabla f(x),\, a(t,x) \rangle + \frac{1}{2} \sum_{i,j=1}^{d} b^{ij}(t,x) \partial_{x^i} \partial_{x^j} f(x),$$

  where $b^{ij}(t,x) = \sigma(t,x)\sigma(t,x)^T$
- Transition density $p(t,x,T,y)$

Forward representation (Feynman Kac formula)

$$u(t,x) = \mathbb{E}\left[f\left(X_{t,x}(T) \right) \right] = \int p(t,x,T,y) f(y) dy =: I(f)$$

$$\partial_t u(t,x) + L_t u(t,x) = 0, \quad u(T,x) = f(x)$$

Reverse process

- Consider: $v(s, y) := \int g(x) p(t, x, s, y) dx$

- Consider: $v(s,y) := \int g(x)p(t,x,s,y)dx$

Fokker-Planck equation:

$$\partial_s p(t,x,s,y) = \frac{1}{2}\sum_{i,j=1}^{d} \partial_{y^i}\partial_{y^j}\left(b^{ij}(s,y)p(t,x,s,y)\right) - \sum_{i=1}^{d} \partial_{y^i}\left(a^i(s,y)p(t,x,s,y)\right)$$

- Consider: $v(s,y) := \int g(x)p(t,x,s,y)dx$

Cauchy problem for v

$$\partial_s v(s,y) = \frac{1}{2}\sum_{i,j=1}^{d} \partial_{y^i}\partial_{y^j}\left(b^{ij}(s,y)v(s,y)\right) - \sum_{i=1}^{d} \partial_{y^i}\left(a^i(s,y)v(s,y)\right),$$

$$v(t,y) = g(y)$$

- Consider: $v(s,y) := \int g(x)p(t,x,s,y)dx$

Cauchy problem for $\widetilde{v}(s,y) := v(T+t-s,y)$

$$\partial_s \widetilde{v}(s,y) + \frac{1}{2}\sum_{i,j=1}^{d} \widetilde{b}^{ij}(s,y)\partial_{y^i}\partial_{y^j}\widetilde{v}(s,y) + \sum_{i=1}^{d} \alpha^i(s,y)\partial_{y^i}\widetilde{v}(s,y) + c(y)\widetilde{v}(s,y) = 0,$$

$$\widetilde{v}(T,y) = g(y),$$

where $\widetilde{b}(s,y) := b(T+t-s,y)$, $\widetilde{a}(s,y) := a(T+t-s,y)$,

$$\alpha^i(s,y) := \sum_{j=1}^{d} \partial_{y^j}\widetilde{b}^{ij}(s,y) - \widetilde{a}^i(s,y),$$

$$c(s,y) = \frac{1}{2}\sum_{i,j=1}^{d} \partial_{y^i}\partial_{y^j}\widetilde{b}^{ij}(s,y) - \sum_{i=1}^{d} \partial_{y^i}\widetilde{a}^i(s,y)$$

- Consider: $v(s,y) := \int g(x)p(t,x,s,y)dx$

Reverse representation (Feynman-Kac formula)

$$I^*(g) := v(T,y) = \mathbb{E}\left[g\left(Y(T)\right)\mathcal{Y}(T)\right]$$
$$dY(s) = \alpha(s,Y(s))ds + \widetilde{\sigma}(s,Y(s))dB(s), \quad Y(t) = y,$$
$$d\mathcal{Y}(s) = c(s,Y(s))\mathcal{Y}(s)ds, \quad \mathcal{Y}(t) = 1$$

$$\alpha^i(s,y) := \sum_{j=1}^{d} \partial_{y^j}\widetilde{b}^{ij}(s,y) - \widetilde{a}^i(s,y),$$
$$c(s,y) = \frac{1}{2}\sum_{i,j=1}^{d} \partial_{y^i}\partial_{y^j}\widetilde{b}^{ij}(s,y) - \sum_{i=1}^{d} \partial_{y^i}\widetilde{a}^i(s,y)$$

Theorem (Milstein, Schoenmakers, Spokoiny 2004)

Choose X and $(Y, \mathcal{Y})$ independent, $t < t^* < T$:

$$\mathbb{E}\left[f\left(X_{t,x}(t^*), Y_{t^*,y}(T)\right)\mathcal{Y}_{t^*,y}(T)\right] =$$
$$= \int p(t, x, t^*, x')f(x', y')p(t^*, y', T, y)dx'dy' =: J(f).$$

- Formally inserting $f(x', y') = \delta_0(x' - y')$ gives $J(f) = p(t, x, T, y)$
- Use kernel $f(x', y') = K_\delta(x' - y') := \delta^{-d}K\left(\frac{x'-y'}{\delta}\right)$ with bandwidth $\delta > 0$

Monte Carlo estimator

$$\hat{p}_{N,M,\delta} := \frac{1}{\delta^d MN}\sum_{n=1}^{N}\sum_{m=1}^{M}\mathcal{Y}^m_{t^*,y}(T)K\left(\frac{X^n_{t,x}(t^*) - Y^m_{t^*,y}(T)}{\delta}\right)$$

Theorem (Milstein, Schoenmakers, Spokoiny 2004)

Choose X and $(Y, \mathcal{Y})$ independent, $t < t^* < T$:

$$\mathbb{E}\left[f\left(X_{t,x}(t^*), Y_{t^*,y}(T)\right)\mathcal{Y}_{t^*,y}(T)\right] =$$
$$= \int p(t,x,t^*,x')f(x',y')p(t^*,y',T,y)dx'dy' =: J(f).$$

- Formally inserting $f(x',y') = \delta_0(x'-y')$ gives $J(f) = p(t,x,T,y)$
- Use kernel $f(x',y') = K_\delta(x'-y') := \delta^{-d}K\left(\frac{x'-y'}{\delta}\right)$ with bandwidth $\delta > 0$

Monte Carlo estimator

$$\hat{p}_{N,M,\delta} := \frac{1}{\delta^d MN}\sum_{n=1}^{N}\sum_{m=1}^{M}\mathcal{Y}^m_{t^*,y}(T)K\left(\frac{X^n_{t,x}(t^*) - Y^m_{t^*,y}(T)}{\delta}\right)$$

Outline

- Consider a grid $0 = s_0 < \cdots < s_{K+L} = T$
- Choose $t^* := s_K$, rename $t_i := s_{i+K}$, $\hat{t}_i := T + t^* - t_{L-i}$, $0 \leq i \leq L$
- Assume $p(s_0, x, T, y) > 0$
- $K : \mathbb{R}^d \to \mathbb{R}$, $\int K(u)du = 1$

Theorem

Let $\hat{t}_i := T + t^* - t_{L-i}$, $X_t = X_{s_0,x}(t)$, $Y_t = Y_{t^*,y}(t)$, $\mathcal{Y}_t = \mathcal{Y}_{t^*,y}(t)$, X *and* $(Y, \mathcal{Y})$ *independent, then*

$$\mathbb{E}\left[g(X_{s_1}, \ldots, X_{s_{K+L-1}}) \,\middle|\, X_T = y \right] = \frac{1}{p(s_0, x, T, y)} \lim_{\delta \downarrow 0} \mathbb{E}\Big[g(X_{s_1}, \ldots, X_{t^*}, Y_{\hat{t}_{L-1}}, \ldots, Y_{\hat{t}_1}) K_\delta \left(Y_T - X_{t^*} \right) \mathcal{Y}_T \Big].$$

- Consider a grid $0 = s_0 < \cdots < s_{K+L} = T$
- Choose $t^* := s_K$, rename $t_i := s_{i+K}$, $\hat{t}_i := T + t^* - t_{L-i}$, $0 \leq i \leq L$
- Assume $p(s_0, x, T, y) > 0$
- $K : \mathbb{R}^d \to \mathbb{R}$, $\int K(u)du = 1$

Theorem

Let $\hat{t}_i := T + t^* - t_{L-i}$, $X_t = X_{s_0,x}(t)$, $Y_t = Y_{t^*,y}(t)$, $\mathcal{Y}_t = \mathcal{Y}_{t^*,y}(t)$, X *and* $(Y, \mathcal{Y})$ *independent, then*

$$\mathbb{E}\left[g(X_{s_1}, \ldots, X_{s_{K+L-1}}) \,\Big|\, X_T = y \right] = \lim_{\delta \downarrow 0} \frac{\mathbb{E}\left[g(X_{s_1}, \ldots, X_{t^*}, Y_{\hat{t}_{L-1}}, \ldots, Y_{\hat{t}_1}) K_\delta (Y_T - X_{t^*}) \mathcal{Y}_T \right]}{\mathbb{E}\left[K_\delta (Y_T - X_{t^*}) \mathcal{Y}_T \right]}.$$

Outline

Conditions on the diffusion process:

Transition densities $p(s,x,t,y)$ and $q(s,x,t,y)$ of X and Y exist and

$$\left|\partial_x^\alpha \partial_y^\beta p(s,x,t,y)\right| \leq \frac{C_1}{(t-s)^\nu} \exp\left(-C_2 \frac{|y-x|^2}{t-s}\right)$$

for multi-indices $|\alpha| + |\beta| \leq r+1$ (and sim. for q). Moreover, $p(s_0, x, T, y) > 0$.

Conditions on the kernel:

- $\int K(v)dv = 1$, $\int v^\alpha K(v)dv = 0$, $0 < |\alpha| \leq r$ (order r kernel)
- $K(v) \leq C \exp\left(-\alpha |v|^{2+\beta}\right)$, $C, \alpha, \beta \geq 0$, $v \in \mathbb{R}^d$

Conditions on the function:

- g and its first and second derivatives are polynomially bounded

- Stochastic representation:

$$H := \mathbb{E}\left[g(X_{s_1}, \ldots, X_{t_{L-1}}) \,\middle|\, X_T = y \right] =$$

$$\lim_{\delta \downarrow 0} \mathbb{E}\Big[g(X_{s_1}, \ldots, X_{t^*}, Y_{\hat{t}_{L-1}}, \ldots, Y_{\hat{t}_1}) K_\delta \left(Y_T - X_{t^*} \right) \mathcal{Y}_T \Big] / p(s_0, x, T, y)$$

- Stochastic representation:

$$H := \mathbb{E}\Big[g(X_{s_1}, \dots, X_{t_{L-1}}) \Big| X_T = y \Big] =$$
$$\lim_{\delta \downarrow 0} \mathbb{E}\Big[g(X_{s_1}, \dots, X_{t^*}, Y_{\hat{t}_{L-1}}, \dots, Y_{\hat{t}_1}) K_\delta (Y_T - X_{t^*}) \mathcal{Y}_T \Big] / p(s_0, x, T, y)$$

- Estimator: for (X^n) i.i.d., $(Y^m, \mathcal{Y}^m)$ i.i.d.,

$$\widehat{H}_{\delta,M,N} := \frac{\sum_{n=1}^N \sum_{m=1}^M g\left(X^n_{s_1}, \dots, X^n_{s_K}, Y^m_{\hat{t}_{L-1}}, \dots, Y^m_{\hat{t}_1} \right) K\left(\frac{Y^m_T - X^n_{t^*}}{\delta} \right) \mathcal{Y}^m_T}{\sum_{n=1}^N \sum_{m=1}^M K\left(\frac{Y^m_T - X^n_{t^*}}{\delta} \right) \mathcal{Y}^m_T} \times$$
$$\times \mathbf{1}_{\frac{1}{NM} \delta^{-d} \sum_{n=1}^N \sum_{m=1}^M K\left(\frac{Y^m_T - X^n_{t^*}}{\delta} \right) \mathcal{Y}^m_T > \overline{p}/2},$$

- Estimator: for (X^n) i.i.d., $(Y^m, \mathcal{Y}^m)$ i.i.d.,

$$\widehat{H}_{\delta,M,N} := \frac{\sum_{n=1}^N \sum_{m=1}^M g\left(X^n_{s_1}, \ldots, X^n_{s_K}, Y^m_{\hat{t}_{L-1}}, \ldots, Y^m_{\hat{t}_1}\right) K\left(\frac{Y^m_T - X^n_{t^*}}{\delta}\right) \mathcal{Y}^m_T}{\sum_{n=1}^N \sum_{m=1}^M K\left(\frac{Y^m_T - X^n_{t^*}}{\delta}\right) \mathcal{Y}^m_T} \times$$

$$\times \mathbf{1}_{\frac{1}{NM}\delta^{-d} \sum_{n=1}^N \sum_{m=1}^M K\left(\frac{Y^m_T - X^n_{t^*}}{\delta}\right) \mathcal{Y}^m_T > \overline{p}/2},$$

- Assume exact simulation of X, Y, $\mathcal{Y}$.

Theorem

Assume $p(s_0, x, T, y) > \overline{p} > 0$, choose $M = N$. Then the mean squared error of the forward-reverse estimator is of order

$$MSE = O\left(\delta^{2(r+1)} + \frac{1}{N} + \frac{1}{N^2\delta^d}\right).$$

$$\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M} K_\delta\left(Y_T^m - X_{t^*}^n\right)\mathcal{Y}_T^m$$

- Assume that kernel K (essentially) has support with diameter r
- Order points $X^n(t^*)$, $Y^m(T)$, $n, m = 1, \dots, N$, into boxes with side length δr (Cost: $\mathcal{O}(N \log N)$)
- Fix $X^n(t^*)$; contribution only from $Y^m(T)$ in the same or neighboring boxes
- Number of such $Y^m(T)$ is $\mathcal{O}(\delta^d N)$ *on average*.
- For $\delta = \mathcal{O}(N^{-1/d})$, average total cost: $\mathcal{O}\left(N \log(N)^2\right)$

$$\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}K_\delta\left(Y_T^m - X_{t^*}^n\right)\mathcal{Y}_T^m$$

- Assume that kernel K_δ (essentially) has support with diameter δr
- Order points $X^n(t^*)$, $Y^m(T)$, $n, m = 1, \ldots, N$, into boxes with side length δr (Cost: $\mathcal{O}(N \log N)$)
- Fix $X^n(t^*)$; contribution only from $Y^m(T)$ in the same or neighboring boxes
- Number of such $Y^m(T)$ is $\mathcal{O}(\delta^d N)$ *on average*.
- For $\delta = \mathcal{O}(N^{-1/d})$, average total cost: $\mathcal{O}\left(N \log(N)^2\right)$

$$\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}K_\delta\left(Y_T^m - X_{t^*}^n\right)\mathcal{Y}_T^m$$

- Assume that kernel K_δ (essentially) has support with diameter δr
- Order points $X^n(t^*)$, $Y^m(T)$, $n, m = 1, \ldots, N$, into boxes with side length δr (Cost: $\mathcal{O}(N \log N)$)
- Fix $X^n(t^*)$; contribution only from $Y^m(T)$ in the same or neighboring boxes
- Number of such $Y^m(T)$ is $\mathcal{O}(\delta^d N)$ *on average*.
- For $\delta = \mathcal{O}(N^{-1/d})$, average total cost: $\mathcal{O}\left(N \log(N)^2\right)$

Implementation of the double sum

$$\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M} K_\delta\left(Y_T^m - X_{t^*}^n\right)\mathcal{Y}_T^m$$

- Assume that kernel K_δ (essentially) has support with diameter δr
- Order points $X^n(t^*)$, $Y^m(T)$, $n, m = 1, \ldots, N$, into boxes with side length δr (Cost: $\mathcal{O}(N \log N)$)
- Fix $X^n(t^*)$; contribution only from $Y^m(T)$ in the same or neighboring boxes
- Number of such $Y^m(T)$ is $\mathcal{O}(\delta^d N)$ *on average*.
- For $\delta = \mathcal{O}(N^{-1/d})$, average total cost: $\mathcal{O}\left(N \log(N)^2\right)$

- Assume that $X_{s_0,x}(s)$, $\left(Y_{t^*,y}(t), \mathcal{Y}_{t^*,y}(T)\right)$ can be simulated exactly at constant cost.
- Cost of simulation step: $\mathcal{O}(N)$
- Cost of "box-ordering" step: $\mathcal{O}(N \log(N))$
- Cost of evaluation step: $\mathcal{O}\left(\max(N^2\delta^d, N)\log(N)^2\right)$

Cost estimate

- Assume that $X_{s_0,x}(s)$, $\left(Y_{t^*,y}(t), \mathcal{Y}_{t^*,y}(T)\right)$ can be simulated exactly at constant cost.
- Cost of simulation step: $\mathcal{O}(N)$
- Cost of "box-ordering" step: $\mathcal{O}(N \log(N))$
- Cost of evaluation step: $\mathcal{O}\left(\max(N^2\delta^d, N)\log(N)^2\right)$

Cost estimate

- Case $d \leq 2(r+1)$: Choose $N = \epsilon^{-2}$, $\delta = \epsilon^{2/d}$, achieve MSE $\mathcal{O}(\epsilon^2)$ at average cost $\mathcal{O}\left(\epsilon^{-2}\log\left(\epsilon^{-1}\right)^2\right)$
- Case $d > 2(r+1)$: Choose $N = \epsilon^{-\left(1+\frac{d}{2(r+1)}\right)}$, $\delta = \epsilon^{1/(r+1)}$, achieve MSE $\mathcal{O}(\epsilon^2)$ at average cost $\mathcal{O}\left(\epsilon^{-\left(1+\frac{d}{2(r+1)}\right)}\log\left(\epsilon^{-1}\right)^2\right)$
- Curse of dimensionality for $d > 2(r+1)$

- Assume that $X_{s_0,x}(s)$, $\left(Y_{t^*,y}(t), \mathcal{Y}_{t^*,y}(T)\right)$ can be simulated exactly at constant cost.
- Cost of simulation step: $\mathcal{O}(N)$
- Cost of "box-ordering" step: $\mathcal{O}(N \log(N))$
- Cost of evaluation step: $\mathcal{O}\left(\max(N^2\delta^d, N)\log(N)^2\right)$

Cost estimate

- Case $d \leq 2(r+1)$: Choose $N = \epsilon^{-2}$, $\delta = \epsilon^{2/d}$, achieve MSE $\mathcal{O}(\epsilon^2)$ at average cost $\mathcal{O}\left(\epsilon^{-2}\log\left(\epsilon^{-1}\right)^2\right)$
- Case $d > 2(r+1)$: Choose $N = \epsilon^{-\left(1+\frac{d}{2(r+1)}\right)}$, $\delta = \epsilon^{1/(r+1)}$, achieve MSE $\mathcal{O}(\epsilon^2)$ at average cost $\mathcal{O}\left(\epsilon^{-\left(1+\frac{d}{2(r+1)}\right)}\log\left(\epsilon^{-1}\right)^2\right)$
- Curse of dimensionality for $d > 2(r+1)$

- Replace exact values $X_{s_0,x}(s)$, $\left(Y_{t^*,y}(t), \mathcal{Y}_{t^*,y}(T)\right)$ by approximated values obtained from Euler discretization
- Use Euler steps of size h

Single level complexity estimate

- Case $d \leq 2(r+1)$: Choose $N = \epsilon^{-2}$, $\delta = \epsilon^{2/d}$, $h = \epsilon$, achieve MSE $\mathcal{O}(\epsilon^2)$ at average cost $\mathcal{O}(\epsilon^{-(2+1)} \log^2 \epsilon^{-1})$
- Case $d > 2(r+1)$: Choose $N = \epsilon^{-\left(1+\frac{d}{2(r+1)}\right)}$, $\delta = \epsilon^{1/(r+1)}$, $h = \epsilon$, achieve MSE $\mathcal{O}(\epsilon^2)$ at average cost $\mathcal{O}\left(\epsilon^{-\left(1+\frac{d}{2(r+1)}+1\right)} \log^2 \epsilon^{-1}\right)$

- Single level estimator for numerator (for $N = M$):

$$\frac{1}{N^2} \sum_{n,m=1}^{N} g_{n,m}^{(h)} K_\delta \left(Y_{T,m}^{(h)} - X_{t^*,n}^{(h)} \right) \mathcal{Y}_{T,m}^{(h)}$$

Multilevel Monte Carlo [B., Nagapetyan, Schoenmakers]

$$\frac{1}{N_0^2}\sum_{n,m=1}^{N_0} g_{n,m}^{(h_0)} K_\delta\left(Y_{T,m}^{(h_0)} - X_{t^*,n}^{(h_0)}\right)\mathcal{Y}_{T,m}^{(h_0)} +$$

$$\sum_{l=1}^{L}\frac{1}{N_l^2}\sum_{n,m=1}^{N_l}\left(g_{n,m}^{(h_l)} K_\delta\left(Y_{T,m}^{(h_l)} - X_{t^*,n}^{(h_l)}\right)\mathcal{Y}_{T,m}^{(h_l)} - g_{n,m}^{(h_{l-1})} K_\delta\left(Y_{T,m}^{(h_{l-1})} - X_{t^*,n}^{(h_{l-1})}\right)\mathcal{Y}_{T,m}^{(h_{l-1})}\right)$$

Multilevel Monte Carlo [B., Nagapetyan, Schoenmakers]

$$\frac{1}{N_0^2}\sum_{n,m=1}^{N_0} g_{n,m}^{(h_0)} K_\delta\left(Y_{T,m}^{(h_0)} - X_{t^*,n}^{(h_0)}\right)\mathcal{Y}_{T,m}^{(h_0)} +$$

$$\sum_{l=1}^{L}\frac{1}{N_l^2}\sum_{n,m=1}^{N_l}\left(g_{n,m}^{(h_l)} K_\delta\left(Y_{T,m}^{(h_l)} - X_{t^*,n}^{(h_l)}\right)\mathcal{Y}_{T,m}^{(h_l)} - g_{n,m}^{(h_{l-1})} K_\delta\left(Y_{T,m}^{(h_{l-1})} - X_{t^*,n}^{(h_{l-1})}\right)\mathcal{Y}_{T,m}^{(h_{l-1})}\right)$$

Some complexity results (ignoring log-terms)

- Euler scheme, $8 < d \leq 2(r+1)$: complexity $\mathcal{O}\left(\epsilon^{-\left(2+\frac{4}{d-2}+1/3\right)}\right)$
- Milstein scheme, $8 < d \leq 2(r+1)$: complexity $\mathcal{O}\left(\epsilon^{-\left(2+\frac{4}{d-2}\right)}\right)$

Further possible multilevel strategies

- Use smaller δ on higher levels: $\delta \to \delta_l$
- Use higher order kernel K on higher levels: $K_\delta \to K_\delta^l$ or $K_\delta \to K_{\delta_l}^l$
- Include "cross-terms": $g_{n,m}^{(h_{l,k})} K_\delta\left(Y_{T,m}^{(h_l)} - X_{t^*,n}^{(h_k)}\right)\mathcal{Y}_{T,m}^{(h_l)}$

Multilevel Monte Carlo [B., Nagapetyan, Schoenmakers]

$$\frac{1}{N_0^2}\sum_{n,m=1}^{N_0} g_{n,m}^{(h_0)} K_\delta\left(Y_{T,m}^{(h_0)} - X_{t^*,n}^{(h_0)}\right)\mathcal{Y}_{T,m}^{(h_0)} +$$

$$\sum_{l=1}^{L}\frac{1}{N_l^2}\sum_{n,m=1}^{N_l}\left(g_{n,m}^{(h_l)} K_\delta\left(Y_{T,m}^{(h_l)} - X_{t^*,n}^{(h_l)}\right)\mathcal{Y}_{T,m}^{(h_l)} - g_{n,m}^{(h_{l-1})} K_\delta\left(Y_{T,m}^{(h_{l-1})} - X_{t^*,n}^{(h_{l-1})}\right)\mathcal{Y}_{T,m}^{(h_{l-1})}\right)$$

Some complexity results (ignoring log-terms)

- Euler scheme, $8 < d \leq 2(r+1)$: complexity $\mathcal{O}\left(\epsilon^{-\left(2+\frac{4}{d-2}+1/3\right)}\right)$
- Milstein scheme, $8 < d \leq 2(r+1)$: complexity $\mathcal{O}\left(\epsilon^{-\left(2+\frac{4}{d-2}\right)}\right)$

Further possible multilevel strategies

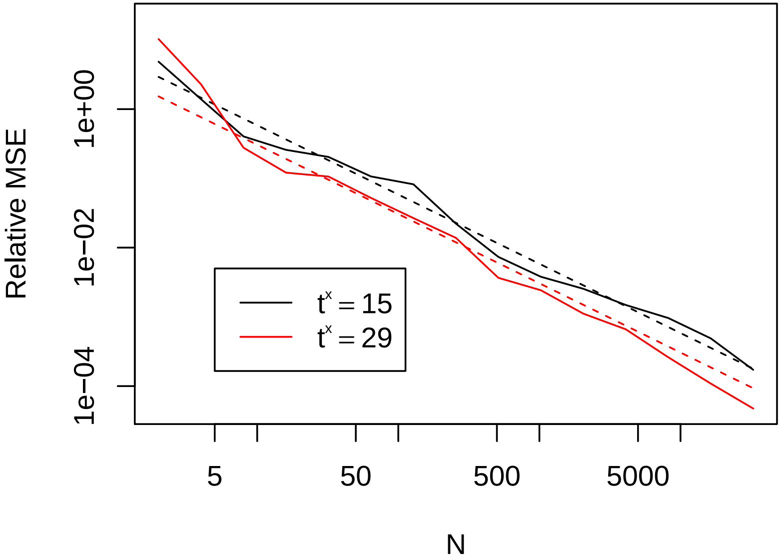
- Use smaller δ on higher levels: $\delta \to \delta_l$
- Use higher order kernel K on higher levels: $K_\delta \to K_\delta^l$ or $K_\delta \to K_{\delta_l}^l$
- Include “cross-terms”: $g_{n,m}^{(h_{l,k})} K_\delta\left(Y_{T,m}^{(h_l)} - X_{t^*,n}^{(h_k)}\right)\mathcal{Y}_{T,m}^{(h_l)}$

- Heston model:

$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dB_t^1,$$
$$dv_t = (\alpha v_t + \beta) dt + \xi \sqrt{v_t} \left(\rho dB_t^1 + \sqrt{1-\rho^2} dB_t^2 \right)$$

- $a(x) = \begin{pmatrix} \mu x_1 \\ \alpha x_2 + \beta \end{pmatrix}, \quad \sigma(x) = \begin{pmatrix} x_1 \sqrt{x_2} & 0 \\ \xi \rho \sqrt{x_2} & \xi \sqrt{1-\rho^2} \sqrt{x_2} \end{pmatrix}$
- Reverse drift: $\alpha(x) = \begin{pmatrix} (2x_2 + \rho\xi - \mu) x_1 \\ (\rho\xi - \alpha) x_2 + \xi^2 - \beta \end{pmatrix}, c(x) = x_2 + \rho\xi - \mu - \alpha.$
- Realized variance: $RV := \sum_{i=1}^{30} \left(\log(S_{t_{i+1}}) - \log(S_{t_i}) \right)^2$
- Objective: $\mathbb{E}[RV \,|\, S_T = s]$, $T = 1/12$

Numerical experiment



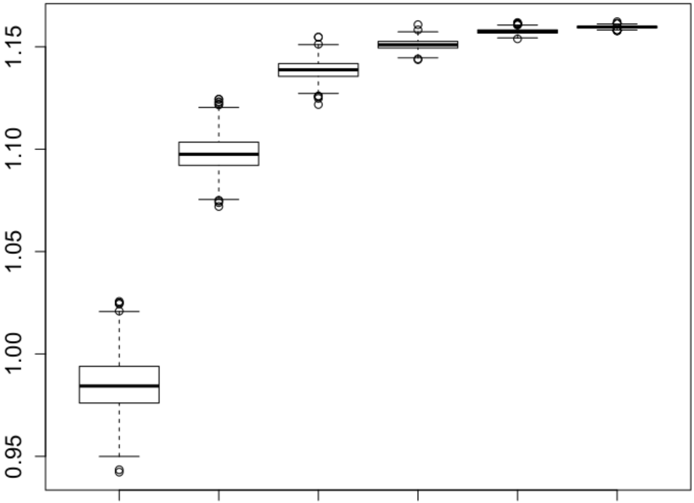
- X a suitable diffusion (or more general Markov) process observed at discrete times $s_0, \ldots, s_K$
- Given data $\mathbf{x} = (X(s_0), \ldots, X(s_K))$, unknown parameter θ
- Log-likelihood function on paths-space $g = g(X;\theta) = g\left((X(t))_{s_0 \le t \le s_K};\theta\right)$

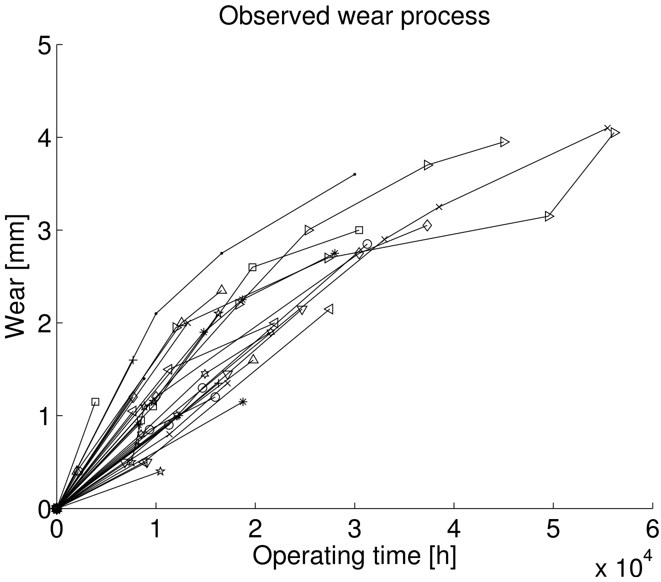
EM algorithm

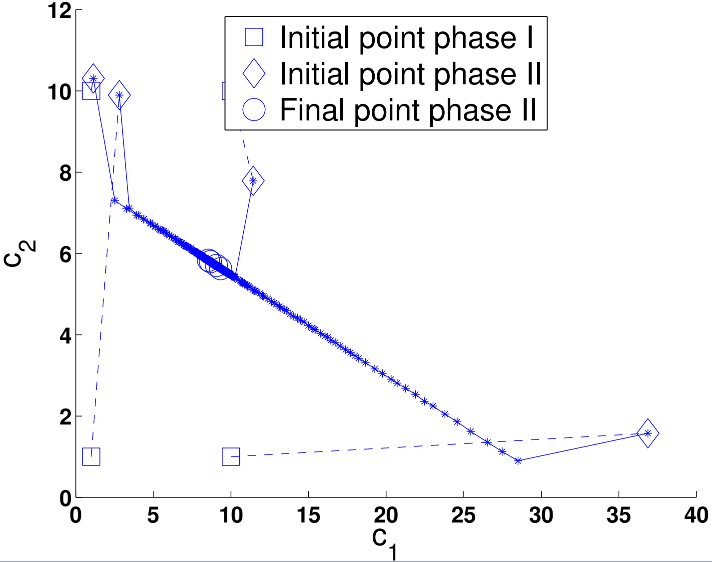
(E) $Q(\theta|\theta_n, \mathbf{x}) := E_{\theta_n}\left[g(X;\theta) \,|\, (X(s_0), \ldots, X(s_K)) = \mathbf{x}\right]$

(M) $\theta_{n+1} := \arg\max_{\theta\in\Theta} Q(\theta|\theta_n, \mathbf{x})$.

EM estimation in a (discrete) OU process [B., Mai, Schoenmakers]




Observed wear process
Wear [mm]
Operating time [h]
x 10^4
0
1
2
3
4
5
6
0
1
2
3
4
5



Initial point phase I
Initial point phase II
Final point phase II
12
10
8
6
4
2
0
0
5
10
15
20
25
30
35
40
c_2
c_1

References

- Bayer, C., Schoenmakers, J.: *Simulation of conditional diffusions via forward-reverse stochastic representations*, AAP, 2014.
- Bayer, C., Mai, H., Schoenmakers, J.: *Forward reverse EM algorithm for Markov chains*, Preprint, 2014.
- Bayer, C., Moraes, A., Tempone, R., Villanova, P.: *The forward-reverse algorithm for stochastic reaction networks with applications to statistical inference*, Preprint 2015.
- Delyon, B., Hu, Y.: *Simulation of conditioned diffusion and application to parameter estimation*, SPA, 2006.
- Dempster, A.P., Laird, N.M., Rubin, D.B.: *Maximum likelihood from incomplete data via the EM algorithm*, J. R. Stat. Soc., 1977.
- Milstein, G., Schoenmakers, J. G. M., Spokoiny, V.: *Transition density estimation for stochastic differential equations via forward–reverse representations*, Bernoulli, 2004.