Simulation of conditional diffusions via forward-reverse stochastic representations

Christian Bayer and John Schoenmakers

SRI UQ Workshop 2015, KAUST
Outline

1. Introduction
2. The forward-reverse method for transition density estimation
3. A forward-reverse representation for conditional expectations
4. The forward-reverse estimator
5. Numerical examples
6. Applications to the EM algorithm
1 Introduction

2 The forward-reverse method for transition density estimation

3 A forward-reverse representation for conditional expectations

4 The forward-reverse estimator

5 Numerical examples

6 Applications to the EM algorithm
Conditional expectations

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

\[
dX(s) = a(s, X(s)) \, ds + \sigma(s, X(s)) \, dB(s), \quad 0 \leq s \leq T
\]

Goal

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

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

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, $0 \leq d' \leq d$. 
Conditional expectations

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

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

Goal

Given a grid $\mathcal{D} = \{ 0 = s_0 < \cdots < s_{K+L+1} = T \}$, $f : \mathbb{R}^{(K+L)d} \rightarrow \mathbb{R}$ 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]
EM algorithm

- 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 $\theta_0$.

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

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

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

- 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 $\theta_0$.

\begin{align*}
(E) & \quad Q(\theta|\theta_n, y) := E_{\theta_n} \left[ \log \left( f(y|U, \theta) h(U; \theta) \right) \mid Y = y \right] \\
(M) & \quad \theta_{n+1} := \arg \max_{\theta \in \Theta} Q(\theta|\theta_n, y).
\end{align*}

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

▶ 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 \( \theta_0 \).

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

(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 \( \nabla l(\theta^*; y) = 0 \)
EM algorithm

- 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 $\theta_0$.

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

(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 $\nabla l(\theta^*; y) = 0$
Drift estimation with the EM algorithm

- **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:** \( x = (x_0, \ldots, x_K) \) of \( X := (X(s_0), \ldots, X(s_K)) \), \( s_0 = 0, s_K = T \)
- **EM algorithm with**
  \[
  Q(\theta|\theta_n, x) = \mathbb{E}_{\theta_n} \left[ -\theta \int_0^T X_s dX_s - \frac{\theta^2}{2} \int_0^T X_s^2 ds \mid X = x \right]
  = \sum_{i=1}^K \mathbb{E}_{\theta_n} \left[ -\theta \int_{s_{i-1}}^{s_i} X_s dX_s - \frac{\theta^2}{2} \int_{s_{i-1}}^{s_i} X_s^2 ds \mid X_{s_{i-1}} = x_{i-1}, X_{s_i} = x_i \right]
  \]
Drift estimation with the EM algorithm

- 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: $x = (x_0, \ldots, x_K)$ of $X := (X(s_0), \ldots, X(s_K))$, $s_0 = 0, s_K = T$

- EM algorithm with

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

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

Forward-reverse for conditional diffusions · January 6, 2015 · Page 6 (28)
1 Introduction

2 The forward-reverse method for transition density estimation

3 A forward-reverse representation for conditional expectations

4 The forward-reverse estimator

5 Numerical examples

6 Applications to the EM algorithm
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)

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

$$\partial_t u(t, x) + L_t u(t, x) = 0, \quad u(T, x) = f(x)$$
Consider: \( v(s, y) := \int g(x)p(t, x, s, y)dx \)
Reverse process

Consider: \( \nu(s, y) := \int g(x)p(t, x, s, y)dx \)

**Fokker-Planck equation:**

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

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 \( \tilde{v}(s, y) := v(T + t - s, y) \)

\[
\partial_s \tilde{v}(s, y) + \frac{1}{2} \sum_{i,j=1}^{d} \tilde{b}^{ij}(s, y) \partial_{y_i} \partial_{y_j} \tilde{v}(s, y) + \sum_{i=1}^{d} \alpha^i(s, y) \partial_{y_i} \tilde{v}(s, y) + c(y)\tilde{v}(s, y) = 0
\]

\[
\tilde{v}(T, y) = g(y),
\]

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

\[
\alpha^i(s, y) := \sum_{j=1}^{d} \partial_{y_j} \tilde{b}^{ij}(s, y) - \tilde{a}^i(s, y),
\]

\[
c(s, y) = \frac{1}{2} \sum_{i,j=1}^{d} \partial_{y_i} \partial_{y_j} \tilde{b}^{ij}(s, y) - \sum_{i=1}^{d} \partial_{y_i} \tilde{a}^i(s, y)
\]
Consider: \( v(s, y) := \int g(x)p(t, x, s, y)dx \)

**Reverse representation (Feynman-Kac formula)**

\[
I^*(g) := v(T, y) = \mathbb{E}[g(Y(T))Y(T)]
\]

\[
dY(s) = \alpha(s, Y(s))ds + \tilde{\sigma}(s, Y(s))dB(s), \quad Y(t) = y,
\]

\[
dY(s) = c(s, Y(s))Y(s)ds, \quad Y(t) = 1
\]

\[
\alpha^i(s, y) := \sum_{j=1}^{d} \partial_y b^{ij}(s, y) - \tilde{a}^i(s, y),
\]

\[
c(s, y) = \frac{1}{2} \sum_{i,j=1}^{d} \partial_y \partial_y b^{ij}(s, y) - \sum_{i=1}^{d} \partial_y \tilde{a}^i(s, y)
\]
**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**

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

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

1 Introduction

2 The forward-reverse method for transition density estimation

3 A forward-reverse representation for conditional expectations

4 The forward-reverse estimator

5 Numerical examples

6 Applications to the EM algorithm
A forward-reverse representation of the conditional expectation

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

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

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

1 Introduction

2 The forward-reverse method for transition density estimation

3 A forward-reverse representation for conditional expectations

4 The forward-reverse estimator

5 Numerical examples

6 Applications to the EM algorithm
Assumptions

**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
The forward-reverse estimator for the conditional expectation

- Stochastic representation:

\[ H := \mathbb{E} \left[ g(X_{s_1}, \ldots, X_{t_L-1}) \mid X_T = y \right] = \]

\[ \lim_{\delta \downarrow 0} \mathbb{E} \left[ g(X_{s_1}, \ldots, X_{t^*}, \hat{Y}_{L-1}, \ldots, \hat{Y}_{t_1}) K_\delta (Y_T - X_{t^*}) \mathcal{Y}_T \right] / p(s_0, x, T, y) \]
The forward-reverse estimator for the conditional expectation

- **Stochastic representation:**

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

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

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

\[
\widehat{H}_{\delta,M,N} := \frac{\sum_{n=1}^{N} \sum_{m=1}^{M} g \left( X_{s_1}^n, \ldots, X_{s_K}^n, Y_{\hat{t}_{L-1}}^m, \ldots, Y_{\hat{t}_1}^m \right) K \left( \frac{Y_T^m - X_{t^*}^n}{\delta} \right) \mathcal{Y}_T^m}{\sum_{n=1}^{N} \sum_{m=1}^{M} K \left( \frac{Y_T^m - X_{t^*}^n}{\delta} \right) \mathcal{Y}_T^m} \times \]

\[
\frac{1}{NM} \delta^{-d} \sum_{n=1}^{N} \sum_{m=1}^{M} K \left( \frac{Y_T^m - X_{t^*}^n}{\delta} \right) \mathcal{Y}_T^m > p/2',
\]

Forward-reverse for conditional diffusions · January 6, 2015 · Page 15 (28)
The forward-reverse estimator for the conditional expectation

- **Estimator:** for \((X^n)\) i.i.d., \((Y^m, Y^m)\) i.i.d.,

\[
\hat{H}_{\delta, M, N} := \frac{\sum_{n=1}^{N} \sum_{m=1}^{M} g\left(X_{s_{1}}, \ldots, X_{s_{K}}, Y_{\hat{t}_{L-1}}, \ldots, Y_{\hat{t}_{1}}\right) K\left(\frac{Y_{T} - X_{n}^*}{\delta}\right) Y_{T}^{m}}{\sum_{n=1}^{N} \sum_{m=1}^{M} K\left(\frac{Y_{T} - X_{n}^*}{\delta}\right) Y_{T}^{m}} \times \prod_{1}^{1} \frac{1}{NM} \delta^{-d} \sum_{n=1}^{N} \sum_{m=1}^{M} K\left(\frac{Y_{T} - X_{n}^*}{\delta}\right) Y_{T}^{m} > \bar{p}/2',
\]

- **Assume exact simulation of** \(X, Y, Y\).

**Theorem**

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

\[
MSE = O\left(\delta^{2(r+1)} + \frac{1}{N} + \frac{1}{N^2 \delta^d}\right).
\]
Implementation of the double sum

\[
\frac{1}{NM} \sum_{n=1}^{N} \sum_{m=1}^{M} K_{\delta} \left( Y_{T}^{m} - X_{t^*}^{n} \right) Y_{T}^{m}
\]

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

\[
\frac{1}{NM} \sum_{n=1}^{N} \sum_{m=1}^{M} K_{\delta} \left( Y_{T}^{m} - X_{t^{*}}^{n} \right) Y_{T}^{m}
\]

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

\[
\frac{1}{NM} \sum_{n=1}^{N} \sum_{m=1}^{M} K_\delta \left( Y^m_T - X^n_t \right) Y^m_T
\]

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

\[ \frac{1}{NM} \sum_{n=1}^{N} \sum_{m=1}^{M} K_{\delta} \left( Y_{T}^{m} - X_{t^*}^{n} \right) Y_{T}^{m} \]

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

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

- Cost of simulation step: $O(N)$
- Cost of “box-ordering” step: $O(N \log(N))$
- Cost of evaluation step: $O\left(\max(N^2 \delta^d, N) \log(N)^2\right)$

### Cost estimate

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

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

Cost estimate

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

- Replace exact values $X_{s_0,x}(s), \left( Y_{t*,y}(t), Y_{t*,y}(T) \right)$ 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 $O(\epsilon^2)$ at average cost $O(\epsilon^{-(2+1)} \log^2 \epsilon^{-1})$
- Case $d > 2(r + 1)$: Choose $N = \epsilon^{-(1 + \frac{d}{2(r+1)})}$, $\delta = \epsilon^{1/(r+1)}$, $h = \epsilon$, achieve MSE $O(\epsilon^2)$ at average cost $O\left(\epsilon^{-(1 + \frac{d}{2(r+1)} + 1)} \log^2 \epsilon^{-1}\right)$
Single level estimator for numerator (for $N = M$):

$$
\frac{1}{N^2} \sum_{n,m=1}^{N} g_{n,m}^{(h)} K_\delta \left( Y_{T,m}^{(h)} - X_{i^*,n}^{(h)} \right) Y_{T,m}^{(h)}
$$
\[
\frac{1}{N_0^2} \sum_{n,m=1}^{N_0} g_{n,m}^{(h_0)} K_\delta \left( Y_{T,m}^{(h_0)} - X_{t^*,n}^{(h_0)} \right) \mathcal{Y}_{T,m}^{(h_0)} + \\
\sum_{l=1}^{L} \frac{1}{N_l^2} \sum_{n,m=1}^{N_l} \left( g_{n,m}^{(h_l)} K_\delta \left( Y_{T,m}^{(h_l)} - X_{t^*,n}^{(h_l)} \right) \mathcal{Y}_{T,m}^{(h_l)} - g_{n,m}^{(h_{l-1})} K_\delta \left( Y_{T,m}^{(h_{l-1})} - X_{t^*,n}^{(h_{l-1})} \right) \mathcal{Y}_{T,m}^{(h_{l-1})} \right)
\]
Multilevel Monte Carlo [B., Nagapetyan, Schoenmakers]

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

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

Some complexity results (ignoring log-terms)

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

Further possible multilevel strategies

- Use smaller \( \delta \) on higher levels: \( \delta \rightarrow \delta_l \)
- Use higher order kernel \( K \) on higher levels: \( K_\delta \rightarrow K^l_\delta \) or \( K_\delta \rightarrow K^l_{\delta_l} \)
- Include “cross-terms”: \( g^{(h_l)}_{n,m} K_\delta \left( Y_{T,m}^{(h_l)} - X_{t^*,n}^{(h_l)} \right) \mathcal{Y}_{T,m}^{(h_l)} \)
Multilevel Monte Carlo [B., Nagapetyan, Schoenmakers]

\[
\frac{1}{N^2_0} \sum_{n,m=1}^{N_0} g^{(h_0)}_{n,m} K_\delta \left( Y^{(h_0)}_{T,m} - X^{(h_0)}_{t^*,n} \right) \mathcal{Y}_{T,m}^{(h_0)} +
\]

\[
\sum_{l=1}^{L} \frac{1}{N^2_l} \sum_{n,m=1}^{N_l} \left( g^{(h_l)}_{n,m} K_\delta \left( Y^{(h_l)}_{T,m} - X^{(h_l)}_{t^*,n} \right) \mathcal{Y}_{T,m}^{(h_l)} - g^{(h_{l-1})}_{n,m} K_\delta \left( Y^{(h_{l-1})}_{T,m} - X^{(h_{l-1})}_{t^*,n} \right) \mathcal{Y}_{T,m}^{(h_{l-1})} \right)
\]

Some complexity results (ignoring log-terms)

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

Further possible multilevel strategies

- Use smaller \(\delta\) on higher levels: \(\delta \rightarrow \delta_l\)
- Use higher order kernel \(K\) on higher levels: \(K_\delta \rightarrow K^l_\delta\) or \(K_\delta \rightarrow K^l_{\delta_l}\)
- Include “cross-terms”: \(g^{(h_{l,k})}_{n,m} K_\delta \left( Y^{(h_l)}_{T,m} - X^{(h_k)}_{t^*,n} \right) \mathcal{Y}_{T,m}^{(h_l)}\)
Outline

1 Introduction

2 The forward-reverse method for transition density estimation

3 A forward-reverse representation for conditional expectations

4 The forward-reverse estimator

5 Numerical examples

6 Applications to the EM algorithm
Conditional expectations of the realized variance

- Heston model:

\[
\begin{align*}
    dS_t &= \mu S_t dt + \sqrt{v_t} S_t dB^1_t, \\
    dv_t &= (\alpha v_t + \beta) dt + \xi \sqrt{v_t} \left( \rho dB^1_t + \sqrt{1 - \rho^2} dB^2_t \right)
\end{align*}
\]

- \( a(x) = \begin{pmatrix} \mu x_1 \\ \alpha x_2 + \beta \end{pmatrix} \), \( \sigma(x) = \begin{pmatrix} x_1 \sqrt{x_2} & 0 \\ \xi \rho \sqrt{x_2} & \xi \sqrt{1 - \rho^2} \sqrt{x_2} \end{pmatrix} \)

- 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: \( RV := \sum_{i=1}^{30} (\log(S_{t_{i+1}}) - \log(S_{t_i}))^2 \)

- Objective: \( \mathbb{E} [RV \mid S_T = s], \quad T = 1/12 \)
Numerical experiment

Relative MSE vs. N

- \( t^* = 15 \)
- \( t^* = 29 \)

Forward-reverse for conditional diffusions · January 6, 2015 · Page 22 (28)
EM again

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

### EM algorithm

**E**
\[
Q(\theta|\theta_n, x) := E_{\theta_n} [g(X; \theta) \mid (X(s_0), \ldots, X(s_K)) = x]
\]

**M**
\[
\theta_{n+1} := \arg \max_{\theta \in \Theta} Q(\theta|\theta_n, x).
\]
EM estimation for wear: Data

Observed wear process

![Graph showing the observed wear process over operating time. The graph plots wear in millimeters (mm) against operating time in hours (h). The x-axis represents operating time from 0 to 6 x 10^4 hours, while the y-axis represents wear from 0 to 5 millimeters. Multiple lines with different markers and styles represent different wear processes.](image-url)
References


