Weierstrass Institute for
Applied Analysis and Stochastics

Leibniz Association

Simulation of conditional diffusions via forward-reverse stochastic representations

Christian Bayer and John Schoenmakers

Linköping, September 11, 2013

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

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

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

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 $A, B \subset \mathbb{R}^d$, compute

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

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

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

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

(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 $\mathbf{X} := (X(s_0), \ldots, X(s_K))$, $s_0 = 0,\ s_K = T$
- Discrete likelihood function in general not available or complicated
- 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: $\mathbf{x} = (x_0, \ldots, x_K)$ of $\mathbf{X} := (X(s_0), \ldots, X(s_K))$, $s_0 = 0,\ s_K = T$
- Discrete likelihood function in general not available or complicated
- 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: $\mathbf{x} = (x_0, \ldots, x_K)$ of $\mathbf{X} := (X(s_0), \ldots, X(s_K))$, $s_0 = 0,\ s_K = T$
- Discrete likelihood function in general not available or complicated
- 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: $\mathbf{x} = (x_0, \ldots, x_K)$ of $\mathbf{X} := (X(s_0), \ldots, X(s_K))$, $s_0 = 0,\ s_K = T$
- Discrete likelihood function in general not available or complicated
- EM algorithm with

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

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

$$dY(s) = \left(a(s, Y(s)) - \frac{Y(s) - y}{T - s}\right) ds + \sigma(s, Y(s)) dB(s), \quad Y(0) = x$$

- SDE admits a unique solution Y on $[0, T[$ with $\lim_{s \to T} Y(s) = y$
- The law of Y on path-space is absolutely continuous w.r.t. the law of X conditioned on $X(T) = y$, $X(0) = x$.
- The Radon-Nikodym derivative is explicitly given (up to a constant) as an integral of $Y(s)$, $\sigma^{-1}(s, Y(s))$ and quadratic co-variations between them.

- Dimension $d = 1$
- $X_t^{(1)}$ solution of SDE started at $X_0^{(1)} = x$
- $X_t^{(2)}$ solution of SDE started at $X_0^{(2)} = y$
- $\tau := \inf\{ 0 \le t \le T \mid X_t^{(1)} = X_{T-t}^{(2)} \}$
- $Z_t := X_t^{(1)}$, $0 \le t \le \tau$, $Z_t := X_{T-t}^{(2)}$, $\tau \le t \le T$ on $\{ \tau \le T \}$.

Theorem (Bladt and Sørensen 2009)

The distribution of Z given $\{ \tau \le T \}$ is equal to the distribution of a bridge process given that the bridge is hit by an independent realization of the SDE with initial distribution $p(T, y, x)dx$.

- Crucial hitting probability of bridge and time-reversed diffusion hard to estimate

- Dimension $d = 1$
- $X_t^{(1)}$ solution of SDE started at $X_0^{(1)} = x$
- $X_t^{(2)}$ solution of SDE started at $X_0^{(2)} = y$
- $\tau := \inf\{0 \leq t \leq T \mid X_t^{(1)} = X_{T-t}^{(2)}\}$
- $Z_t := X_t^{(1)}, 0 \leq t \leq \tau, Z_t := X_{T-t}^{(2)}, \tau \leq t \leq T$ on $\{\tau \leq T\}$.

Theorem (Bladt and Sørensen 2009)

The distribution of Z given $\{\tau \leq T\}$ is equal to the distribution of a bridge process given that the bridge is hit by an independent realization of the SDE with initial distribution $p(T, y, x)dx$.

- Crucial hitting probability of bridge and time-reversed diffusion hard to estimate

- Dimension $d = 1$
- $X_t^{(1)}$ solution of SDE started at $X_0^{(1)} = x$
- $X_t^{(2)}$ solution of SDE started at $X_0^{(2)} = y$
- $\tau := \inf \{ 0 \leq t \leq T \mid X_t^{(1)} = X_{T-t}^{(2)} \}$
- $Z_t := X_t^{(1)}$, $0 \leq t \leq \tau$, $Z_t := X_{T-t}^{(2)}$, $\tau \leq t \leq T$ on $\{ \tau \leq T \}$.

Theorem (Bladt and Sørensen 2009)

The distribution of Z given $\{ \tau \leq T \}$ is equal to the distribution of a bridge process given that the bridge is hit by an independent realization of the SDE with initial distribution $p(T, y, x)dx$.

- Crucial hitting probability of bridge and time-reversed diffusion hard to estimate

- 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 $b^{ij}(x) = \sigma(t,x)\sigma(t,x)^T$

- Transition density $p(t,x,T,y)$

Forward representation (Feynman Kac formula)

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

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

- 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 $b^{ij}(x) = \sigma(t,x)\sigma(t,x)^T$

- Transition density $p(t,x,T,y)$

Forward representation (Feynman Kac formula)

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

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

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

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

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

where

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

- 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

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

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

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

Reverse representation (Feynman-Kac formula)

$$I^*(g) := v(T,y) = \mathbb{E}\left[g\left(Y_{t,y}(T)\right)\mathcal{Y}_{t,y,1}(T)\right]$$

$$dY(s) = \alpha(s,Y(s))ds + \widetilde{\sigma}(s,Y(s))dB(s), \quad Y(t) = y,$$

$$d\mathcal{Y}(s) = c(s,Y(s))\mathcal{Y}(s)ds, \quad \mathcal{Y}(t) = 1$$

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

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

Theorem (Milstein, Schoenmakers, Spokoiny 2004)

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

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

Proof.

- Condition on $X_{t,x}(t^*)$ and apply the reverse representation
- Integrate with respect to the law of $X_{t,x}(t^*)$

□

Theorem (Milstein, Schoenmakers, Spokoiny 2004)

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

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

Proof.

- Condition on $X_{t,x}(t^*)$ and apply the reverse representation
- Integrate with respect to the law of $X_{t,x}(t^*)$

□

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

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

Theorem (Milstein, Schoenmakers, Spokoiny 2004)

Assume that the coefficients of the SDE are C^∞ bounded and satisfy a uniform ellipticity (or uniform Hörmander) condition.

- *If $d \leq 4$, choose $M = N$, $\epsilon_N = CN^{-1/4}$, then the MSE of $\hat{p}_{N,N,\epsilon_N}$ is of order N^{-1}.*
- *For $d > 4$, choose $M = N$ and $\epsilon_N = CN^{-2/(4+d)}$, then the MSE of $\hat{p}_{N,N,\epsilon_N}$ is of order $N^{-8/(4+d)}$.*

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

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

Theorem (Milstein, Schoenmakers, Spokoiny 2004)

Assume that the coefficients of the SDE are C^∞ bounded and satisfy a uniform ellipticity (or uniform Hörmander) condition.

- *If $d \le 4$, choose $M = N$, $\epsilon_N = CN^{-1/4}$, then the MSE of $\hat{p}_{N,N,\epsilon_N}$ is of order N^{-1}.*
- *For $d > 4$, choose $M = N$ and $\epsilon_N = CN^{-2/(4+d)}$, then the MSE of $\hat{p}_{N,N,\epsilon_N}$ is of order $N^{-8/(4+d)}$.*

Theorem

Introduce the re-ordered grid $t^* < \hat{t}_1 < \cdots < \hat{t}_L = T$ *defined by* $\hat{t}_i := T + t^* - t_{L-i}$, $i = 1, \ldots, L$. *Then*

$$\mathbb{E}\left[f(Y_{t^*,y}(T), Y_{t^*,y}(\hat{t}_{L-1}), \ldots, Y_{t^*,y}(\hat{t}_1))\mathcal{Y}_{t^*,y}(T)\right] = \int_{\mathbb{R}^{d\times L}} f(y_1, y_2, ..., y_L) \prod_{i=1}^{L} p(t_{i-1}, y_i, t_i, y_{i+1}) dy_i.$$

Proof.

- $Y_{t^*,y}(s) = Y_{0,y}(s - t^*) =: Y_{y;T}(s - t^*)$,
  $\mathcal{Y}_{t^*,y}(s) = \mathcal{Y}_{0,y}(s - t^*) =: \mathcal{Y}_y(s - t^*)$
- $\mathbb{E}\left[f(Y_{y;T}(T - t^*))\mathcal{Y}_{y;T}(T - t^*)\right] = \int p(t^*, y', T, y) f(y') dy'$
- Induction in L. □

Theorem

Introduce the re-ordered grid $t^* < \hat{t}_1 < \cdots < \hat{t}_L = T$ *defined by* $\hat{t}_i := T + t^* - t_{L-i}$, $i = 1, \ldots, L$. *Then*

$$\mathbb{E}\left[f(Y_{t^*,y}(T), Y_{t^*,y}(\hat{t}_{L-1}), \ldots, Y_{t^*,y}(\hat{t}_1))\mathcal{Y}_{t^*,y}(T)\right] =$$

$$\int_{\mathbb{R}^{d\times L}} f(y_1, y_2, ..., y_L) \prod_{i=1}^{L} p(t_{i-1}, y_i, t_i, y_{i+1}) dy_i.$$

Proof.

- $Y_{t^*,y}(s) = Y_{0,y}(s - t^*) =: Y_{y;T}(s - t^*)$,
  $\mathcal{Y}_{t^*,y}(s) = \mathcal{Y}_{0,y}(s - t^*) =: \mathcal{Y}_y(s - t^*)$
- $\mathbb{E}\left[f(Y_{y;T}(T - t^*))\mathcal{Y}_{y;T}(T - t^*)\right] = \int p(t^*, y', T, y) f(y') dy'$
- Induction in L.

□

Theorem

Introduce the re-ordered grid $t^* < \hat{t}_1 < \cdots < \hat{t}_L = T$ *defined by* $\hat{t}_i := T + t^* - t_{L-i}$, $i = 1, \ldots, L$. *Then*

$$\mathbb{E}\left[f(Y_{t^*,y}(T), Y_{t^*,y}(\hat{t}_{L-1}), \ldots, Y_{t^*,y}(\hat{t}_1))\mathcal{Y}_{t^*,y}(T)\right] = \int_{\mathbb{R}^{d\times L}} f(y_1, y_2, ..., y_L) \prod_{i=1}^{L} p(t_{i-1}, y_i, t_i, y_{i+1}) dy_i.$$

Proof.

- $Y_{t^*,y}(s) = Y_{0,y}(s - t^*) =: Y_{y;T}(s - t^*)$,
  $\mathcal{Y}_{t^*,y}(s) = \mathcal{Y}_{0,y}(s - t^*) =: \mathcal{Y}_y(s - t^*)$
- $\mathbb{E}\left[f(Y_{y;T}(T - t^*))\mathcal{Y}_{y;T}(T - t^*)\right] = \int p(t^*, y', T, y) f(y') dy'$
- Induction in L.

□

- Consider a grid $0 = s_0 < \cdots < s_{K+L} = T$
- Choose $t^* := s_K$, rename $t_i := s_{i+K}$, $0 \leq i \leq L$
- 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*

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

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 $|\alpha| + |\beta| \leq 2$ (and sim. for q). Moreover, $p(s_0,x,T,y) > 0$.

Conditions on the kernel:

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

Conditions on the function:

- g and its first and second derivatives are polynomially bounded

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 $|\alpha| + |\beta| \leq 2$ (and sim. for q). Moreover, $p(s_0, x, T, y) > 0$.

Conditions on the kernel:

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

Conditions on the function:

- g and its first and second derivatives are polynomially bounded

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 $|\alpha| + |\beta| \leq 2$ (and sim. for q). Moreover, $p(s_0, x, T, y) > 0$.

Conditions on the kernel:

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

Conditions on the function:

- g and its first and second derivatives are polynomially bounded

- Stochastic representation:

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

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

- Stochastic representation:

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

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

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

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

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

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

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

Theorem

Assume $p(s_0, x, T, y) > \overline{p} > 0$, choose $M = N$.

- *For $d \leq 4$, set $\epsilon_N = CN^{-1/4}$, then $\mathbb{E}\left[\left(H - \widehat{H}_{N,N,\epsilon_N}\right)^2\right] = O(N^{-1})$.*
- *For $d > 4$, set $\epsilon_N = CN^{-2/(4+d)}$, then $\mathbb{E}\left[\left(H - \widehat{H}_{N,N,\epsilon_N}\right)^2\right] = O(N^{-8/(4+d)})$.*

- Assume K has compact support in $B_r(0)$

Algorithm

1. *Simulate N indep. trajectories $(X^n)_{n=1}^N$ started at $X_{s_0} = x$ and $(Y^m)_{m=1}^N$ started at $Y_{t^*} = y$ on $\mathcal{D} \cap [s_0, t^*]$ and $\widehat{\mathcal{D}} \cap [t^*, T]$, resp.*
2. *For fixed $m \in \{1, \dots, N\}$, find the sub-sample*

$$\{X_{s_0,x}^{n_k^m}(t^*) : k = 1, \dots, l_m\} := \{X_{s_0,x}^n(t^*) : n = 1, \dots, N\} \cap B_{r\epsilon}(Y_{t^*,y}^m)$$

3. *Evaluate*

$$\widehat{H}_{\epsilon,M,N} \leftarrow \frac{\sum_{m=1}^N \sum_{k=1}^{l_m} g\left(X_{s_1}^{n_k^m}, \dots, X_{s_K}^{n_k^m}, Y_{\hat{t}_{L-1}}^m, \dots, Y_{\hat{t}_1}^m\right) K\left(\frac{Y_T^m - X_{t^*}^{n_k^m}}{\epsilon}\right) \mathcal{Y}_T^m}{\sum_{m=1}^N \sum_{k=1}^{l_m} K\left(\frac{Y_T^m - X_{t^*}^{n_k^m}}{\epsilon}\right) \mathcal{Y}_T^m}$$

- Assume K has compact support in $B_r(0)$

Algorithm

1. *Simulate* N *indep. trajectories* $(X^n)_{n=1}^N$ *started at* $X_{s_0} = x$ *and* $(Y^m)_{m=1}^N$ *started at* $Y_{t^*} = y$ *on* $\mathcal{D} \cap [s_0, t^*]$ *and* $\widehat{\mathcal{D}} \cap [t^*, T]$*, resp.*
2. *For fixed* $m \in \{1, \ldots, N\}$*, find the sub-sample*

$$\{X_{s_0,x}^{n_k^m}(t^*) : \; k = 1, \ldots, l_m\} := \{X_{s_0,x}^n(t^*) : \; n = 1, \ldots, N\} \cap B_{r\epsilon}(Y_{t^*,y}^m)$$

3. *Evaluate*

$$\widehat{H}_{\epsilon,M,N} \leftarrow \frac{\sum_{m=1}^N \sum_{k=1}^{l_m} g\left(X_{s_1}^{n_k^m}, \ldots, X_{s_K}^{n_k^m}, Y_{\hat{t}_{L-1}}^m, \ldots, Y_{\hat{t}_1}^m\right) K\left(\frac{Y_T^m - X_{t^*}^{n_k^m}}{\epsilon}\right) \mathcal{Y}_T^m}{\sum_{m=1}^N \sum_{k=1}^{l_m} K\left(\frac{Y_T^m - X_{t^*}^{n_k^m}}{\epsilon}\right) \mathcal{Y}_T^m}$$

- 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: $O(N)$
- Cost of "box-ordering" step: $O(N\log(N))$ (up to comparisons of integers)
- Cost of evaluation step: $O\left(N^2\epsilon^d\right)$

Complexity estimate

- Case $d \leq 4$: Choose $\epsilon = (N/\log N)^{-1/d}$, achieve MSE $O(N^{-1})$ at cost $O(N\log N)$
- Case $d > 4$: Choose $\epsilon = N^{-2/(4+d)}$, achieve MSE $O(N^{-8/(4+d)})$ at cost $O(N\log N)$

- 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: $\mathcal{O}(N\log(N))$ (up to comparisons of integers)
- Cost of evaluation step: $\mathcal{O}\left(N^2\epsilon^d\right)$

Complexity estimate

- Case $d \leq 4$: Choose $\epsilon = (N/\log N)^{-1/d}$, achieve MSE $\mathcal{O}(N^{-1})$ at cost $\mathcal{O}(N\log N)$
- Case $d > 4$: Choose $\epsilon = N^{-2/(4+d)}$, achieve MSE $\mathcal{O}(N^{-8/(4+d)})$ at cost $\mathcal{O}(N\log N)$

- 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: $\mathcal{O}(N \log(N))$ (up to comparisons of integers)
- Cost of evaluation step: $\mathcal{O}\left(N^2 \epsilon^d\right)$

Complexity estimate

- Case $d \leq 4$: Choose $\epsilon = (N/\log N)^{-1/d}$, achieve MSE $\mathcal{O}(N^{-1})$ at cost $\mathcal{O}(N \log N)$
- Case $d > 4$: Choose $\epsilon = N^{-2/(4+d)}$, achieve MSE $\mathcal{O}(N^{-8/(4+d)})$ at cost $\mathcal{O}(N \log N)$

- 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: $RV := \sum_{i=1}^{30} (\log(S_{t_{i+1}}) - \log(S_{t_i}))^2$
- Objective: $\mathbb{E}[RV \,|\, S_T = s], T = 1/12$

- Heston model:

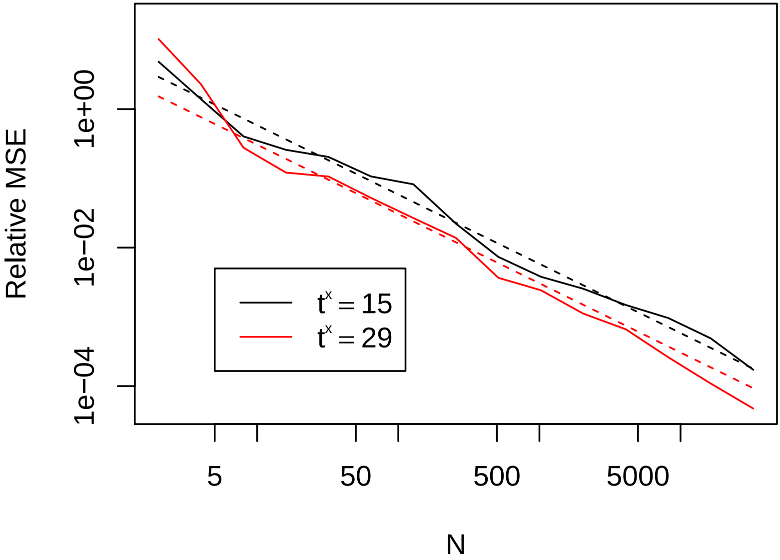
$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dB_t^1,$$
$$dv_t = (\alpha v_t + \beta)dt + \xi \sqrt{v_t}\left(\rho dB_t^1 + \sqrt{1-\rho^2} dB_t^2\right)$$

- $a(x) = \begin{pmatrix} \mu x_1 \\ \alpha x_2 + \beta \end{pmatrix}, \quad \sigma(x) = \begin{pmatrix} x_1 \sqrt{x_2} & 0 \\ \xi\rho \sqrt{x_2} & \xi \sqrt{1-\rho^2} \sqrt{x_2} \end{pmatrix}$
- 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: $\mathbb{E}\left[RV \mid S_T = s\right], T = 1/12$

- 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: $\mathbb{E}\left[RV \mid S_T = s\right]$, $T = 1/12$


Relative MSE
1e+00
1e−02
1e−04
5
50
500
5000
N
t^x = 15
t^x = 29

- Bayer, C., Schoenmakers, J.: Simulation of conditional diffusions via forward-reverse stochastic representations, preprint 2013.
- Bladt, M., Sørensen, M.: *Simple simulation of diffusion bridges with application to likelihood inference for diffusions*, preprint 2012.
- Delyon, B., Hu, Y.: *Simulation of conditioned diffusion and application to parameter estimation*, SPA 116(11), 1660–1675, 2006.
- Kusuoka, S., Stroock, D.: *Applications of the Malliavin calculus, Part II*, J. Fac. Sci. Univ. Yokyo Sec. 1A 32, 1–76, 1985.
- Milstein, G., Schoenmakers, J. G. M., Spokoiny, V.: *Transition density estimation for stochastic differential equations via forward–reverse representations*, Bernoulli 10(2), 281–312, 2004.