Parameter estimation by contrast minimization for noisy discrete observations of a diffusion process

Benjamin Favetto
http://www.favetto.net/

MAP5 - Université Paris Descartes

June 19, 2010

Dynstoch meeting - Angers 2010

1. Model and Assumptions
   - Model and Assumptions
2. Statistical estimation by contrast minimization
   - The Euler-like contrasts
   - Limit theorems for functionals of the observed process
   - Associated central limit theorems
   - Minimum contrast estimators
3. Examples and numerical results
   - The Ornstein-Uhlenbeck process
     - Simulation study
     - Neuronal data
   - The Cox-Ingersoll-Ross process
4. Concluding remarks

Model and Assumptions

Consider

$$\begin{aligned} dX_t &= b(X_t,\kappa)dt + \sigma(X_t,\lambda)dB_t, \quad X_0 = \eta && \text{(hidden)} \\ Y_{i\delta_N} &= X_{i\delta_N} + \rho_N \varepsilon_{i\delta_N} && \text{(observed)} \end{aligned}$$

where

- $\theta = (\kappa, \lambda) \in \Theta_1 \times \Theta_2 = \Theta$ a product of compact intervals
- (B_t) is a Brownian motion and η independent of (B_t)
- the discretization step $\delta_N \to 0$ and the number of observations $N \to \infty$ with $N\delta_N \to \infty$
- the SDE admits an unique (strong) solution (X_t)
- $(\varepsilon_{i\delta_N})$ is a sequence of i.i.d. centered r.v. with $\mathbb{E}((\varepsilon_{i\delta_N})^2) = 1$
- Aim : estimate $\theta_0 = (\kappa_0, \lambda_0) \in \mathring{\Theta}$

Assumptions on the hidden diffusion

(A1) Functions $b(.) = b(., \theta_0)$ and $\sigma(.) = \sigma(., \theta_0)$ belong to $\mathcal{C}^2(\mathbb{R})$ with linear growth,

(A2) The diffusion (X_t) admits a stationary probability $\nu_0(dx) = \nu_0(x)dx$.

(A3) $\forall k > 0$, ν_0 admits finite moment of order k.

(A4) $\forall k > 0$, $\sup_{t \geq 0} \mathbb{E}(|X_t|^k) < \infty$.

Identifiability condition

$$\begin{aligned} \sigma(x, \lambda) = \sigma(x, \lambda_0) &\qquad \nu_0 \text{ almost everywhere implies } \lambda = \lambda_0, \\ b(x, \kappa) = b(x, \kappa_0) &\qquad \nu_0 \text{ almost everywhere implies } \kappa = \kappa_0. \end{aligned}$$

Assumptions on the noise

ρ_N is assumed to be known in the sequel.

(B0) The common distribution of the random variables $\varepsilon_{i\delta_N}$ admits a 8th order moment, and is symmetric.

(B1) $\rho_N = \rho > 0$.

(B2) $\rho_N \to 0$ when $N \to \infty$.

(B2) corresponds to the case $\varepsilon_{i\delta_N} = V_{(i+1)\delta_N} - V_{i\delta_N}$ with (V_t) Brownian motion independent of (B_t).

Minimum contrast estimation for a discretely observed diffusion

The Euler Scheme

Consider (X_t) the solution of $dX_t = b(X_t, \kappa)dt + \sigma(X_t, \lambda)dB_t$ and $\delta_N > 0$. For $\delta_N \to 0$,

$$X_{(j+1)\delta_N} - X_{j\delta_N} \approx \mathcal{N}(b(X_{j\delta_N}, \kappa)\delta_N, \sigma(X_{j\delta_N}, \lambda)^2\delta_N).$$

Contrast for $(X_{j\delta_N})$ based on Gaussian loglikelihood

$$\mathcal{C}_N(\theta) = \sum_{j=0}^{N-1} \left\{ \frac{(X_{(j+1)\delta_N} - X_{j\delta_N} - \delta_N b(X_{j\delta_N}, \kappa))^2}{\delta_N c(X_{j\delta_N}, \lambda)} + \log(c(X_{j\delta_N}, \lambda)) \right\}$$

where $c(., \lambda) = \sigma(., \lambda)^2$, with $N \to \infty$ and $N\delta_N \to \infty$.
$\hat{\theta}_N = (\hat{\kappa}_N, \hat{\lambda}_N) = \arg\min_{\theta \in \Theta} \mathcal{C}_N(\theta)$ is a consistent estimator ([Kessler, 1997])

Local means of the observations

Local means and noise reduction

Consider p_N, k_N such that $p_N = \delta_N^{-\frac{1}{\alpha}}$ for $1 < \alpha \leq 2$, $N = p_N k_N$ and let $\Delta_N = p_N \delta_N = \delta_N^{1-\frac{1}{\alpha}}$. Hence $N\delta_N = k_N \Delta_N$. Define

$$\begin{aligned} Y_\bullet^j &= \frac{1}{p_N}\sum_{i=0}^{p_N-1} Y_{j\Delta_N + i\delta_N} \\ &= \frac{1}{p_N}\sum_{i=0}^{p_N-1} X_{j\Delta_N + i\delta_N} + \frac{1}{p_N}\sum_{i=0}^{p_N-1} \rho_N \varepsilon_{j\Delta_N + i\delta_N} \\ &= X_\bullet^j + \rho_N \varepsilon_\bullet^j \end{aligned}$$

Idea : $Y_\bullet^j \approx X_\bullet^j \approx X_{j\Delta_N} \rightarrow$ Build a contrast function based on $(Y_\bullet^j)$in the ergodic case ([Gloter and Jacod, 2001] for a fixed-length interval)

Minimum contrasts estimation for an hidden diffusion

Contrast for $(Y_\bullet^j)$

$$\mathcal{E}_N(\theta) = \sum_{j=1}^{k_N-2} \left\{ \frac{3}{2} \frac{(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1}, \kappa))^2}{\Delta_N c(Y_\bullet^{j-1}, \lambda)} + \log(c(Y_\bullet^{j-1}, \lambda)) \right\}$$

where $c(., \lambda) = \sigma(., \lambda)^2$.

Modified contrast for $(Y_\bullet^j)$

$$\mathcal{E}_N^{\rho_N}(\theta) = \sum_{j=1}^{k_N-2} \left\{ \frac{3}{2} \frac{(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1}, \kappa))^2}{\Delta_N c_{N,\rho_N}(Y_\bullet^{j-1}, \lambda)} + \log(c_{N,\rho_N}(Y_\bullet^{j-1}, \lambda)) \right\}$$

where $c_{N,\rho_N}(x, \lambda) = \sigma(x, \lambda)^2 + 3\Delta_N^{\frac{2-\alpha}{\alpha-1}} \rho_N^2$, for $1 < \alpha \leq 2$.

1. Model and Assumptions
   - Model and Assumptions
2. Statistical estimation by contrast minimization
   - The Euler-like contrasts
   - Limit theorems for functionals of the observed process
   - Associated central limit theorems
   - Minimum contrast estimators
3. Examples and numerical results
   - The Ornstein-Uhlenbeck process
     - Simulation study
     - Neuronal data
   - The Cox-Ingersoll-Ross process
4. Concluding remarks

Ergodic theorem for local means

For f a function of linear growth uniformly in θ, we have

$$\bar{\nu}_N(f(.,\theta)) = \frac{1}{k_N} \sum_{j=0}^{k_N-1} f(Y_\bullet^j,\theta) \longrightarrow \nu_0(f(.,\theta))$$

uniformly in θ, in probability, as $N \to \infty$, with $\delta_N \to 0$, $p_N \to \infty$, $k_N = N\delta_N^{\frac{1}{\alpha}} \to \infty$, $\Delta_N \to 0$ and $N\delta_N \to \infty$.

Variation of the local means

For f a function of linear growth uniformly in θ, with $p_N = \delta_N^{-\frac{1}{\alpha}}$, $\alpha \in (1,2]$, we have

$$\bar{I}_N(f(.,\theta)) = \frac{1}{k_N\Delta_N} \sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1},\theta)(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1},\kappa)) \xrightarrow{\mathbb{P}} 0$$

uniformly in θ, as $N \to \infty$, with $\delta_N \to 0$, $p_N \to \infty$, $k_N \to \infty$, $\Delta_N \to 0$ and $N\delta_N = k_N\Delta_N \to \infty$.

Quadratic variation of the local means

Define $\bar{Q}_N(f(.,\theta)) = \frac{1}{k_N\Delta_N}\sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1},\theta)(Y_\bullet^{j+1}-Y_\bullet^j)^2$ and assume $p_N = \delta_N^{-\frac{1}{\alpha}}$.

1. If **(B1)** ($\rho_N = \rho > 0$), then $\bar{Q}_N(f(.,\theta)) \xrightarrow{\mathbb{P}} \frac{2}{3}\nu_0(f(.,\theta)\sigma^2(.)) + 2\rho^2\mathbf{1}_{\{\alpha=2\}}\nu_0(f(.,\theta))$,
2. If **(B2)** ($\rho_N \to 0$), then $\bar{Q}_N(f(.,\theta)) \xrightarrow{\mathbb{P}} \frac{2}{3}\nu_0(f(.,\theta)\sigma^2)$

where all the convergences in probability are uniform in $\theta \in \Theta$, as $N \to \infty$, with $\delta_N \to 0$, $p_N \to \infty$, $k_N \to \infty$, $\Delta_N \to 0$ and $N\delta_N = k_N\Delta_N \to \infty$.

Central Limit Theorems for $\bar{I}_N(f)$ and $\bar{Q}_N(f)$ (I)

We do not consider any dependence in θ for f. To prove the asymptotic normality of our estimators, and deal with confidence intervals, we need

CLT for the variation

Assume $\alpha \in (1, 2]$ and $N\delta_N^{3-\frac{2}{\alpha}} \to 0$, we have

$$\sqrt{N\delta_N}\bar{I}_N(f) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu_0(f^2\sigma^2)).$$

Remark : the convergence rate is $\sqrt{N\delta_N}$, which is the usual rate for directly observed diffusions in the drift parameter estimation.

Central Limit Theorems for $\bar{I}_N(f)$ and $\bar{Q}_N(f)$ (II)

CLT for the quadratic variation

Assume $N\delta_N^{2-\frac{1}{\alpha}} \to 0$. Then, with $N\delta_N^{\frac{1}{\alpha}} = k_N \to \infty$, we have

$$\sqrt{N\delta_N^{\frac{1}{\alpha}}}(\bar{Q}_N(f) - \bar{\nu}_N(f(\frac{2}{3}\sigma^2 + 2\rho_N^2\Delta_N^{\frac{2-\alpha}{\alpha-1}})) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V_0(f))$$

where

- if $\alpha \in (1,2)$ and **(B1)** or if $\alpha \in (1,2]$ and **(B2)**

$$V_0(f) = \nu_0(f(.)^2\sigma(.)^2),$$

- if $\alpha = 2$ and $\rho_N = \rho > 0$ **(B1)**, we have

$$V_0(f) = \nu_0(f(.)^2(\sigma(.)^4 + 4\sigma(.)^2\rho^2 + 12\rho^4)).$$

Minimum contrast estimators (I)

We consider $\hat{\theta}_N = \underset{\theta\in\Theta}{\arg\inf}\, \mathcal{E}_N(\theta)$ and $\hat{\theta}_N^{\rho_N} = \underset{\theta\in\Theta}{\arg\inf}\, \mathcal{E}_N^{\rho_N}(\theta)$.

Consistency

Assume that b,c,c^{-1} and their partial derivatives have a linear growth.

1. If **(B1/2)** holds, with $p_N = \delta_N^{-\frac{1}{\alpha}}$ with $\alpha \in (1, 2)$, the estimator $\hat{\theta}_N$ is consistent.
2. If **(B1/2)** holds, with $p_N = \delta_N^{-\frac{1}{\alpha}}$, $\alpha \in (1, 2]$, the estimator $\hat{\theta}_N^{\rho_N}$ is consistent.

Minimum contrast estimators (II)

Asymptotic normality

Assume that $N\delta_N^{2-\frac{1}{\alpha}} \to 0$, when $N \to \infty$, $\delta_N \to 0$, $N\delta_N \to \infty$, $k_N = N\delta_N^{\frac{1}{\alpha}} \to \infty$, $\Delta_N = \delta_N^{1-\frac{1}{\alpha}} \to 0$. If $\rho_N = \rho$ **(B1)** and $\alpha \in (1,2)$ or $\rho_N \to 0$ **(B2)** with $\alpha \in (1,2]$,

$$\begin{pmatrix} \sqrt{N\delta_N}(\hat{\kappa}_N^{\rho_N} - \kappa_0) \\ \sqrt{N\delta_N^{\frac{1}{\alpha}}}(\hat{\lambda}_N^{\rho_N} - \lambda_0) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \mathbf{V}(\theta_0)\right)$$

where

$$\mathbf{V}(\theta_0) = \begin{pmatrix} \left\{\nu_0\left(\frac{(\partial_\kappa b(.,\kappa_0))^2}{c(.,\lambda_0)}\right)\right\}^{-1} & 0 \\ 0 & \frac{9}{4}\left\{\nu_0\left(\frac{(\partial_\lambda c(.,\lambda_0))^2}{c(.,\lambda_0)^2}\right)\right\}^{-1} \end{pmatrix}.$$

In the case $\alpha = 2$ with **(B1)** ($\rho_N = \rho > 0$), the asymptotic variance $\mathbf{V}(\theta_0)$ is increased by the noise variance ρ^2.

The hidden Ornstein-Uhlenbeck model on simulations (I)

The model

$$
\begin{aligned}
dX_t &= \kappa X_t dt + \lambda dB_t, \quad X_0 = x_0 \in \mathbb{R}, \\
Y_{i\delta_N} &= X_{i\delta_N} + \rho\varepsilon_i, \quad \varepsilon_i \sim_{iid} \mathcal{N}(0,1).
\end{aligned}
$$

with $\kappa < 0$ and $\lambda > 0$.

The estimators

$$
\begin{aligned}
\hat{\lambda}_N^2 &= \frac{3}{2k_N\Delta_N} \sum_{j=1}^{k_N-2} (Y_\bullet^{j+1} - Y_\bullet^{j} - \Delta_N \hat{\kappa}_N Y_\bullet^{j-1})^2; \\
\hat{\kappa}_N &= \frac{1}{\Delta_N} \frac{\sum_{j=1}^{k_N-2} Y_\bullet^{j-1}(Y_\bullet^{j+1} - Y_\bullet^{j})}{\sum_{j=1}^{k_N-2} (Y_\bullet^{j-1})^2}.
\end{aligned}
$$

The hidden Ornstein-Uhlenbeck model on simulations (II)

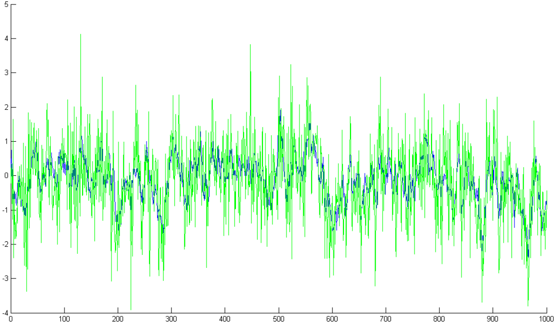


FIG.: This figure represents an Ornstein-Uhlenbeck process with high-frequency noisy observations.
$N = 1000, \delta = 0.1, \kappa = -1, \sigma^2 = 1, \rho^2 = 1$

Influence of the length of the Δ-blocks ($\kappa_0 = -1, \lambda_0 = 1, \rho^2 = 1$)

$N = 5000, \delta = 0.01$ ($N\delta = 50, N\delta^2 = 0.5$) 500 replications

	$\alpha = 1.17 (p = 50, k = 100)$	$\alpha = 1.5 (p = 22, k = 227)$	$\alpha = 2 (p = 10, k = 500)$
$\hat{\kappa}_N$ (10^2 Var)	-0.58 (1.53)	-0.76 (2.75)	-0.82 (3.26)
$\hat{\lambda}^2_N$ (10^2 Var)	0.76 (1.19)	1.07 (1.25)	0.86 (2.61)

TAB.: Influence of the length of the Δ-blocks (N=5000 obs.)

$N = 20000, \delta = 0.005$ ($N\delta = 100, N\delta^2 = 0.5$) 500 replications

	$\alpha = 1.35 (p = 50, k = 400)$	$\alpha = 1.5 (p = 34, k = 588)$	$\alpha = 2 (p = 14, k = 1428)$
$\hat{\kappa}_N$ (10^2 Var)	-0.74 (1.08)	-0.81 (1.47)	-0.87 (1.51)
$\hat{\lambda}^2_N$ (10^3 Var)	0.95 (3.87)	1.05 (3.88)	0.92 (11.07)

TAB.: Influence of the length of the Δ-blocks (N=20000 obs.)

$N = 100000, \delta = 0.001$ ($N\delta = 100, N\delta^2 = 0.1$) 500 replications

	$\alpha = 1.3 (p = 200, k = 500)$	$\alpha = 1.5 (p = 100, k = 1000)$	$\alpha = 2 (p = 32, k = 3125)$
$\hat{\kappa}_N$ (10^2 Var)	-0.81 (1.36)	-0.89 (1.49)	-0.96 (1.95)
$\hat{\lambda}^2_N$ (10^3 Var)	0.90 (2.74)	1.02 (1.99)	0.92 (3.85)

TAB.: Influence of the length of the Δ-blocks (N=100000 obs.)

Influence of other factors

$N = 10^5, \delta = 10^{-3}, \alpha = 1.5, \kappa_0 = -1, \lambda_0 = 1$

	$\rho^2 = 0.1$	$\rho^2 = 1$	$\rho^2 = 2$	$\rho^2 = 5$
$\hat{\kappa}_N$ (10^2 Var)	-0.91 (1.49)	-0.89 (1.50)	-0.86 (1.75)	-0.83 (1.52)
$\hat{\lambda}^2_N$ (10^3 Var)	0.96 (1.71)	1.17 (2.92)	1.47 (4.33)	2.37 (13.42)

TAB.: Influence of ρ^2

$N = 10^5, \delta = 10^{-3}, \alpha = 1.5, \kappa_0 = -1, \rho^2 = 1$

	$\lambda^2 = 0.1$	$\lambda^2 = 0.5$	$\lambda^2 = 1$	$\lambda^2 = 2$
$\hat{\kappa}_N$ (10^2 Var)	-0.81 (1.48)	-0.87 (1.54)	-0.90 (1.64)	-0.89 (1.62)
$\hat{\lambda}^2_N$ (10^3 Var)	0.23 (0.12)	0.58 (0.78)	1.01 (1.95)	2.01 (6.93)

TAB.: Influence of λ^2

$N = 10^5, \delta = 10^{-3}, \alpha = 1.5, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$

	$\mathcal{N}(0,1)$	Laplace$(0, \frac{1}{\sqrt{2}})$	Uniform$(-\sqrt{3}, \sqrt{3})$	Logistic$(0, \frac{\sqrt{3}}{\pi})$
$\hat{\kappa}_N$ (10^2 Var)	-0.89 (1.65)	-0.90 (1.52)	-0.87 (1.53)	-0.89 (1.65)
$\hat{\lambda}^2_N$ (10^3 Var)	1.02 (2.11)	1.02 (2.18)	1.31 (3.45)	1.02 (2.10)

TAB.: Influence of the distribution of ε

The hidden Ornstein-Uhlenbeck model on neuronal data

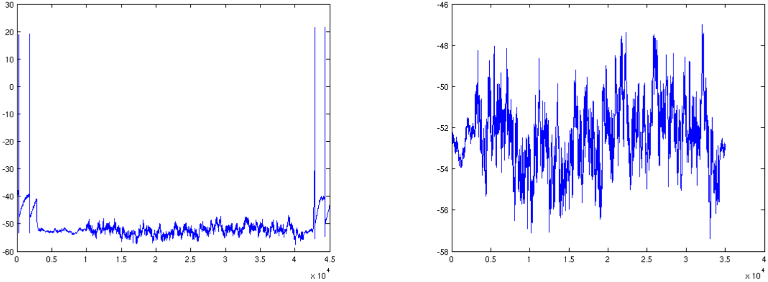


FIG.: Neuronal data ($N = 35000$ observations, $\delta = 0.0002$)

Ornstein-Uhlenbeck model for neuronal data

The model

$$\begin{aligned} dX_t &= (-\frac{X_t}{\tau} + \kappa)dt + \lambda dB_t, \quad X_0 = x_0 \in \mathbb{R}, \\ Y_{i\delta_N} &= X_{i\delta_N} + \rho\varepsilon_i, \quad \varepsilon_i \sim_{iid} \mathcal{N}(0,1). \end{aligned}.$$

with $\hat{\rho}_N^2 = 0.0014$ on this dataset.

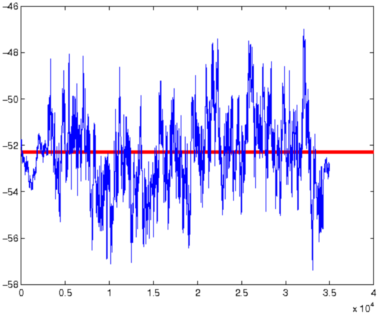
	$p = 23(\alpha = 1.25)$	$p = 14(\alpha = 1.5)$	$p = 7(\alpha = 2)$
$\hat{\tau}_N$ (10^{-2} seconds)	13.52	5.85	4.67
$\hat{\kappa}_N$ (10^2 mV/ sec)	-3.87	-8.95	-11.23
$\hat{\lambda}_N^2$	1.94	1.78	1.10

TAB.: Parameter estimation for neuronal data (with error measurement).

	$\tilde{\tau}_N$ (10^{-2} seconds)	$\tilde{\kappa}_N$ (10^2 mV/ sec)	$\tilde{\lambda}_N^2$
$N = 35000, \delta = 0.0002$	40	-1.31	0.14

TAB.: Parameter estimation for neuronal data (without error measurement).

Mean of the stationary distribution



Notice that the mean of the stationary distribution is given by $\mu = \tau\kappa$, and is usually well estimed. We find

- $\hat{\mu}_N = \hat{\tau}_N \hat{\kappa}_N = -52.28mV$ for the estimator based on the noisy observations model,
- $\tilde{\mu}_N = \tilde{\tau}_N \tilde{\kappa}_N = -52.45mV$ for the estimator based on the direct observations.

1. Model and Assumptions
   - Model and Assumptions
2. Statistical estimation by contrast minimization
   - The Euler-like contrasts
   - Limit theorems for functionals of the observed process
   - Associated central limit theorems
   - Minimum contrast estimators
3. Examples and numerical results
   - The Ornstein-Uhlenbeck process
     - Simulation study
     - Neuronal data
   - The Cox-Ingersoll-Ross process
4. Concluding remarks

The hidden Cox Ingersoll Ross model on simulations

The model

$$\begin{aligned} dX_t &= (\kappa X_t + \kappa')dt + \lambda\sqrt{X_t}dB_t, \\ Y_{t_i} &= X_{t_i}\exp(\varepsilon_{t_i}) \end{aligned}$$

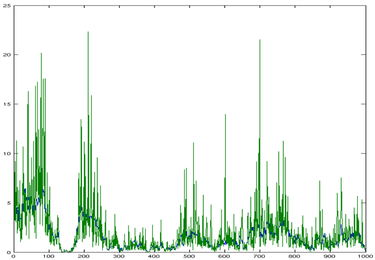
where (ε_{t_i}) is a sequence of independent $\mathcal{N}(0, \rho^2)$ random variables.

Define $Z_t = \log(X_t)$ and Z_t solves

$$dZ_t = (\kappa + (\kappa' - \frac{\lambda^2}{2})\exp(-Z_t))dt + \lambda\exp(-\frac{Z_t}{2})dB_t.$$

Set $\kappa'' = \kappa' - \frac{\lambda^2}{2}$ and then we find explicit estimators.

Results



$\kappa_0 = -2, \kappa_0'' = 1, \lambda_0 = 4, \rho^2 = 0.5, \alpha = 1.5$

	$N = 5.10^3, \delta = 10^{-2}$	$N = 2.10^4, \delta = 5.10^{-3}$	$N = 10^5, \delta = 10^{-3}$
$\hat{\kappa}_N$ (10^2 Var)	-1.43 (6.28)	-1.56 (3.14)	-1.78 (3.37)
$\hat{\kappa}''_N$ (10^2 Var)	0.99 (4.57)	1.03 (2.12)	1.13 (2.44)
$\hat{\lambda}^2_N$ (10^2 Var)	4.23 (37.61)	4.35 (15.15)	4.40 (8.15)

TAB.: Parameter estimation for the CIR diffusion

Concluding remarks

- More general structure of observations : $Y_{i\delta_N} = F(X_{i\delta_N}, dy)$
- Integrated diffusion process plus noise :

$$X_{\bullet}^{j} \approx \frac{1}{\Delta_N} \int_{j\Delta_N}^{(j+1)\Delta_N} X_s ds$$

- Other type of asymptotics

Concluding remarks

- More general structure of observations : $Y_{i\delta_N} = F(X_{i\delta_N}, dy)$
- Integrated diffusion process plus noise :

$$X_{\bullet}^{j} \approx \frac{1}{\Delta_N} \int_{j\Delta_N}^{(j+1)\Delta_N} X_s ds$$

- Other type of asymptotics

Thanks !

References I

Gloter, A. and Jacod, J. (2001).
Diffusions with measurement errors. II. Optimal estimators.
ESAIM Probab. Statist., 5 :243–260 (electronic).

Kessler, M. (1997).
Estimation of an ergodic diffusion from discrete observations.
Scand. J. Statist., 24(2) :211–229.