Statistical research of models geherated by discretely observed solution to SDE's

Ivanenko Dmytro

Kyiv National Taras Shevchenko University, Ukraine

The talk partially contains results obtained in a collaboration with
A. Kulik, A. Kohatsu-Higa, S. Bodnarchuk, R. Pogorielov and Yu. Scatarenko

Potsdam, 25.03 – 03.04.2019

Introduction

The talk is divided into two parts:

The first one is devoted to research asymptotic properties of estimators for discretely observed solution to Lévy driven SDE's

$$dX_t = a_\theta(X_t)dt + dZ_t,$$

where Z is a Lévy process without diffusion component.

The following tasks are investigated:

- Estimation of the Fischer information.
- Effectiveness of the estimation method.
- One-Step estimator.
- Software implementation of the algorithms.

We consider one parameter case and multi parameters one.

Second part

We consider the statistical model generated by SDE

$$dX_t = a(\alpha; X_t)dt + \sigma(\beta; X_t)dW_t.$$

Here, W is a one dimensional Wiener process and the coefficients a and σ are differentiable functions with bounded derivatives.

- Estimation of the transition probability density.
- Construct QMLE and One-Step estimator.
- Software implementation.

Estimation of the Fisher information

We consider the Markov process given by equation

$$\mathrm{d}X_t = a_\theta(X_t)\mathrm{d}t + \mathrm{d}Z_t.$$

Let Z be a Lévy process without a diffusion component that is

$$Z_t = ct + \int_0^t \int_{|u|>1} u\nu(\mathrm{d}s, \mathrm{d}u) + \int_0^t \int_{|u|\le 1} u\tilde{\nu}(\mathrm{d}s, \mathrm{d}u),$$

where ν is a Poisson point measure with the intensity measure $\mathrm{d}s\mu(\mathrm{d}u)$, and $\tilde{\nu}(\mathrm{d}s, \mathrm{d}u) = \nu(\mathrm{d}s, \mathrm{d}u) - \mathrm{d}s\mu(\mathrm{d}u)$ is respective compensated Poisson measure.

We assume the Lévy measure μ to satisfy the following:

(i) for some $\kappa > 0$,

$$\int_{|u|\geq 1} u^{2+\kappa}\mu(du) < \infty;$$

(ii) for some $u_0 > 0$, the restriction of μ on $[-u_0, u_0]$ has a positive density $\sigma \in C^2([-u_0, 0) \cup (0, u_0])$;

(iii) there exists C_0 such that

$$|\sigma'(u)| \leq C_0|u|^{-1}\sigma(u), \quad |\sigma''(u)| \leq C_0 u^{-2}\sigma(u), \quad |u| \in (0, u_0];$$

(iv)

$$\left(\log\frac{1}{\varepsilon}\right)^{-1}\mu\Big(\{u : |u| \geq \varepsilon\}\Big) \to \infty, \quad \varepsilon \to 0.$$

Assumptions about drift

Assume the following:

(i) Let a has bounded derivatives $\partial^{i+j}_{x^i\theta^j} a$, $i \le 3, j \le 2$

(ii) Derivatives $\partial_x a$, $\partial^2_{xx} a$, $\partial^2_{x\theta} a$, $\partial^3_{xxx} a$, $\partial^3_{xx\theta} a$, $\partial^3_{x\theta\theta} a$, $\partial^4_{xxx\theta} a$ are bounded and

$$|a_\theta(x)| + |\partial_\theta a_\theta(x)| + |\partial^2_{\theta\theta} a_\theta(x)| \le C(1 + |x|).$$

for all $\theta \in \Theta,\ x \in R$

(iii) for given $\theta_0 \in \Theta$ there is a neighborhood $(\theta_-, \theta_+) \subset \Theta$ of the point θ_0, such that

$$\limsup_{|x| \to +\infty} \frac{a_\theta(x)}{x} < 0 \quad \text{uniformly w.r.t. } \theta \in (\theta_-, \theta_+).$$

Theorem 1. Representation for the transition probability density and its logarithmic derivative

- X has a transition probability density p_t^θ w.r.t. the Lebesgue measure, which has an integral representation

$$p_t^\theta(x,y) = \mathrm{E}_x^\theta\left[\Xi_t I_{X_t>y}\right], \quad t>0, \quad x,y\in\mathrm{R}.$$

$$\Xi_t := \delta\left(\frac{1}{\mathrm{D}X_t}\right) = \frac{\delta(1)}{\mathrm{D}X_t} + \frac{\mathrm{D}^2X_t}{(\mathrm{D}X_t)^2}.$$

- p_t^θ has a derivative $\partial_\theta p_t^\theta(x,y) = g_t^\theta(x,y)p_t^\theta(x,y)$ with

$$g_t^\theta(x,y) = \begin{cases} \mathrm{E}_{x,y}^{t,\theta}\Xi_t^1, & p_t^\theta(x,y)>0, \\ 0, & \text{otherwise.}\end{cases}$$

$$\Xi_t^1 := \delta\left(\frac{\partial_\theta X_t}{\mathrm{D}X_t}\right) = \frac{(\partial_\theta X_t)\delta(1)}{\mathrm{D}X_t} + \frac{(\partial_\theta X_t)\mathrm{D}^2X_t}{(\mathrm{D}X_t)^2} - \frac{\mathrm{D}(\partial_\theta X_t)}{\mathrm{D}X_t}.$$

Fischer information

The logarithm of the transition probability density has a continuous derivative w.r.t. θ on the open subset of $(0, \infty) \times \mathrm{R} \times \mathrm{R} \times \Theta$ and, on this subset, admits the integral representation

$$\partial_\theta \log p_t^\theta(x, y) = g_t^\theta(x, y) = \mathrm{E}_{x,y}^{t,\theta} \Xi_t^1.$$

Let $x \in \mathrm{R}$, $n \in \mathrm{N}$, $0 < t_1 < \cdots < t_n$ be fixed. Respective Fischer information equals to

$$I(\theta) = \sum_{k=1}^{n} \mathrm{E}_x^\theta \Big(g_{t_k - t_{k-1}}^\theta (X_{t_{k-1}}, X_{t_k}) \Big)^2,$$

Effectiveness checking algorithm in one-parametric model is:

Step 1: Select the method of estimation and build an estimator $\hat{\theta}_n$ of the unknown parameter θ_0

Step 2: Generate N trajectories of the process X with $\hat{\theta}_n = \theta_0$ and for each of them construct a sample size n

Step 3: Calculate estimators $\hat{\theta}_n^k, k = 1, \dots, N$ and find a sample variance $s_N^2 = \frac{1}{N}\sum_{k=1}^N (\sqrt{n}(\hat{\theta}_n^k - \theta)^2)$

Step 4: Calculate n_0 and generate N trajectories of the process X with $\hat{\theta}_n = \theta_0$, for each of them calculate $\Xi_n^k(n_0), k = 1, \dots, N$

Step 5: Find a sample mean $J_n(\hat{\theta}_n, n_0) = \frac{1}{N}\sum_{k=1}^N (\Xi_n^k(n_0))^2$

Step 6: By value $\sqrt{J_n(\hat{\theta}_n, n_0)s_N^2}$ decide on the effectiveness of the method.

Simultaneous equations to find Ξ_t^1 items

Let $Y_t^1 = \partial_\theta X_t, Y_t^2 = DX_t, Y_t^3 = D\partial_\theta X_t, Y_t^4 = D^2 X_t$.

Than $\overline{Y}_t := (Y_t^1, Y_t^2, Y_t^3, Y_t^4)$ is the solution to:

$$\begin{cases} dY_t^1 = \partial_x a_\theta(X_t) Y_t^1 dt + \partial_\theta a_\theta(X_t) dt; \\ dY_t^2 = \partial_x a_\theta(X_t) Y_t^2 dt + d\mathrm{D}Z_t; \\ dY_t^3 = \partial_x a_\theta(X_t) Y_t^3 dt + \left(\partial_{x\theta} a_\theta(X_t) Y_t^2 + \partial_{xx} a_\theta(X_t) Y_t^1 Y_t^2\right) dt; \\ dY_t^4 = \partial_x a_\theta(X_t) Y_t^4 dt + \partial_{xx} a_\theta(X_t)(Y_t^2)^2 dt + d\mathrm{D}^2 Z_t; \\ Y_{t_0}^i = 0, \; i = 1, \ldots, 4, \end{cases}$$

where

$$\mathrm{D}Z_t = \int_0^t \int_{\mathrm{R}} \varrho(u) \nu(\mathrm{d}s, \mathrm{d}u), \quad \mathrm{D}^2 Z_t = \int_0^t \int_{\mathrm{R}} \varrho(u) \varrho'(u) \nu(\mathrm{d}s, \mathrm{d}u),$$

$$\delta(1) = -\int_0^t \int_{\mathrm{R}} \frac{(\sigma(u)\varrho(u))'}{\sigma(u)} \tilde{\nu}(\mathrm{d}s, \mathrm{d}u) \quad \varrho(u) = \begin{cases} u^2, & |u| \le u_1; \\ 0, & |u| \ge u_0. \end{cases}$$

Citation

Expression of Fisher's information was presented in

- Ivanenko D.O., Kulik A.M. (2015) Malliavin calculus approach to statistical inference for Levy driven SDE's.
  Meth. and Comp. in Appl. Prob., vol. 17, p. 107–123.

The effectiveness checking algorithm has been described in

- Ivanenko D.O., Bodnarchuk S.V. (2015) Efficiency of estimating method, in statistical models controlled by Levy noise.
  Th. of Prob. and Math. Stat., vol. 92, p. 9 – 22.

To construct the One-step estimator we need to expression of the second derivative. The following result was presented in

- Ivanenko D.O. (2014) Second derivative of the log-likelihood in the model given by a Lévy driven SDE's.
  Visn. Kyiv Nat. Univ., arxiv org, 1410.2880, vol. 2, p. 18 –22.

Example 1.

Figure: Efficiency in one-parameter nonlinear model.

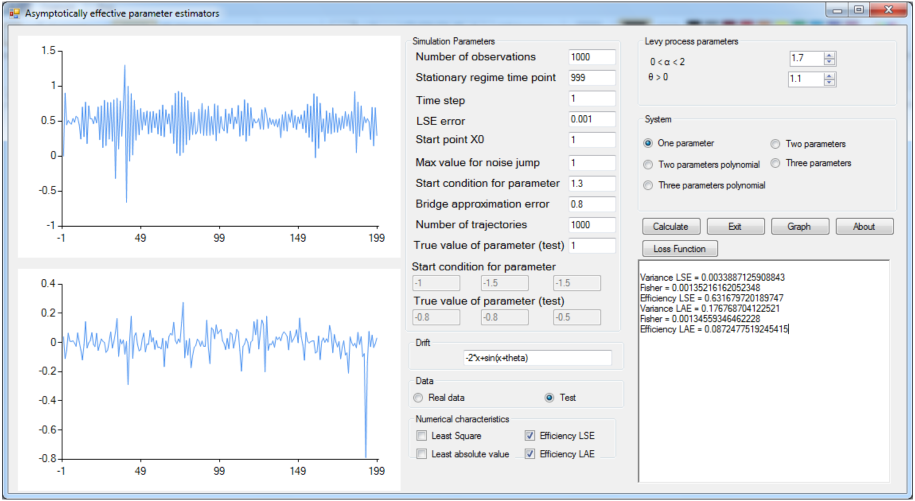
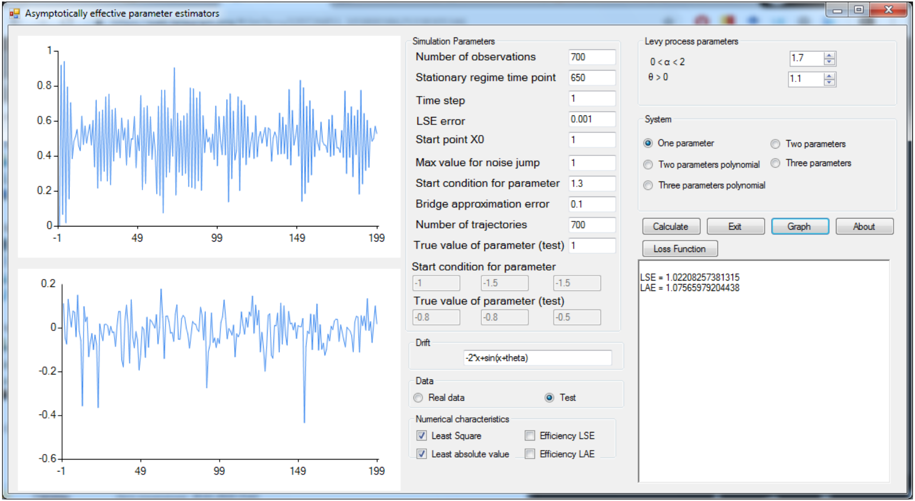


Figure: LSE and LAE in one-parameter nonlinear model.



Example 2.

Figure: Efficiency in two-parameter nonlinear model.

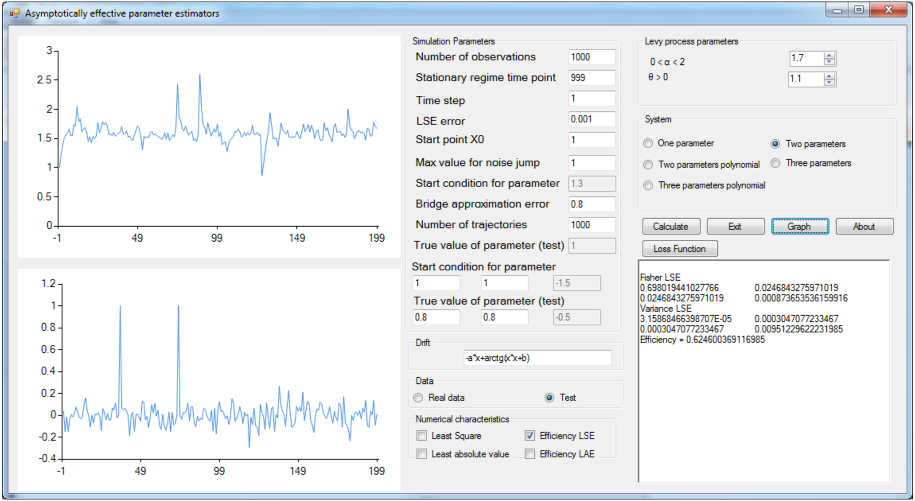
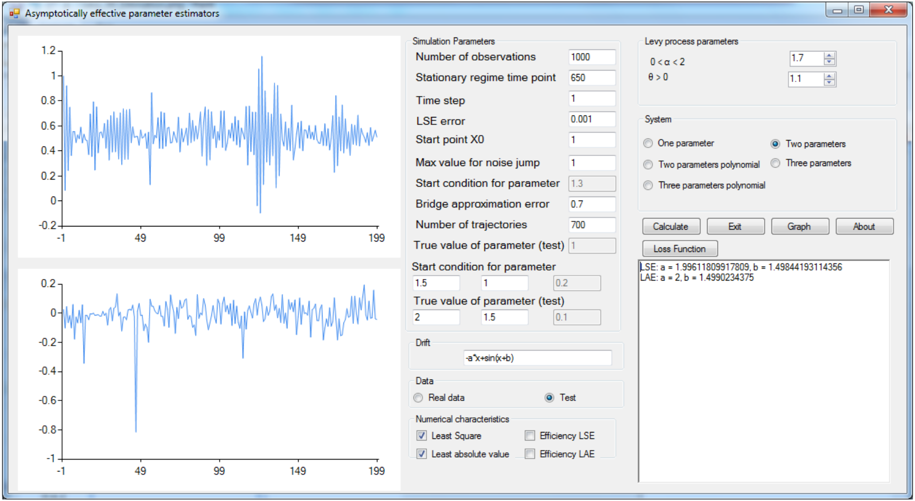


Figure: LSE and LAE in two-parameters nonlinear model.



Example 3.

Figure: Efficiency in three-parameter nonlinear model.

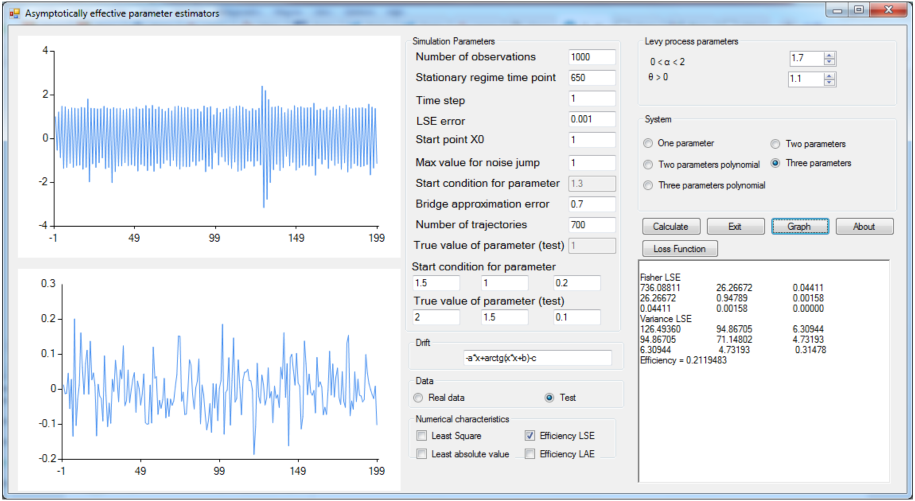
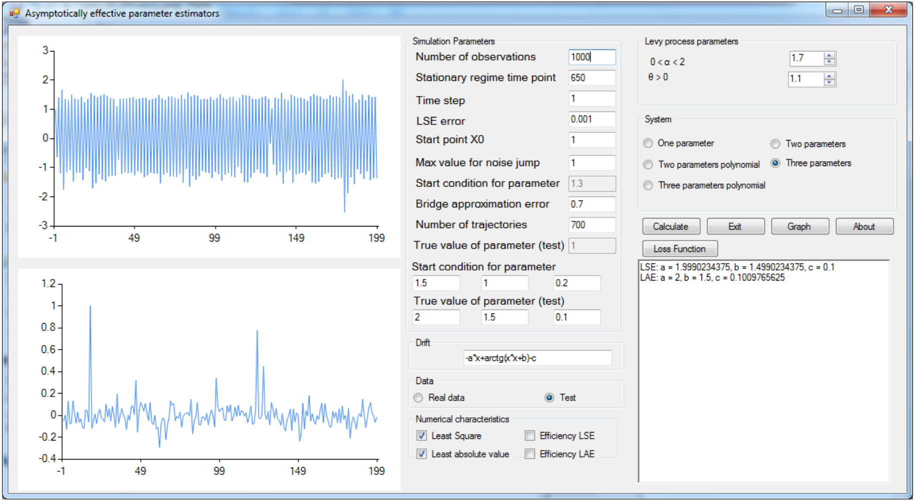


Figure: LSE and LAE in three-parameter nonlinear model.



One-Step estimator

To produce a consistent estimator with asymptotic variance equal to the inverse Fisher information as this is the best possible variance we can achieve for consistent estimators.

- Newton's Method: $\theta_0^{NM} = \theta^{LSE}$ or (like in example) θ^{Lp}, $p = 3$;

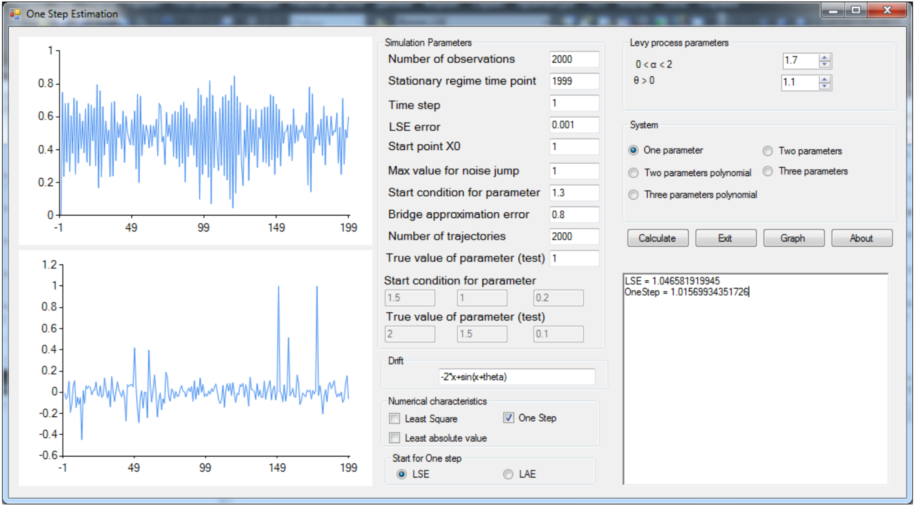
$$\theta_{k+1}^{NM} = \theta_k^{NM} - \left(\partial_{\theta\theta}^2 l(\theta_k^{NM})\right)^{-1} \partial_\theta l(\theta_k^{NM}).$$

- Method of Scoring (Rao's accumulation method): $\theta_0^{NM} = \theta^{LSE}$,

$$\theta_{k+1}^{SM} = \theta_k^{SM} + \left(I(\theta_k^{SM})\right)^{-1} \partial_\theta l(\theta_k^{SM}).$$

Example 4.

Figure: One-Step estimator nonlinear model.



Second part

Consider the stochastic differential equation of the form

$$dX_t = a(\alpha; X_t)dt + \sigma(\beta; X_t)dW_t,$$

where $X_0 = x_0$ and $\theta = (\alpha, \beta) \in \Theta \subseteq R^2$ are unknown parameters, W is a one dimensional Wiener process and the coefficients a and σ are differentiable functions with bounded derivatives.

The following tasks are investigated:

- Develop a software to obtain a decomposition of the transition probability density in degrees t and in terms of Hermit functions.
- Use the previous decomposition to construct QMLE and One-Step estimators.

Density decomposition

We say that the function g is of order at least r if there exists some positive C and c such that

$$|g_t(\theta, x, y)| \leq Ct^r \phi_{ct}(x, y),$$

ϕ is the transition probability density at y for the Wiener process at time t which starts at x.

We want to obtain a decomposition of transition probability density of X in the form

$$p = \sum_{k,m} c_{k,m} t^k \Upsilon^{(m)}(a', b') + r,$$

where r is the function of fixed order at least n and Υ is the Hermite function given by ($m \geq 0$, H_m is the Hermite polynomial)

$$\Upsilon_t^{(m)}(a, b, x, y) = H_m(y - x - at, bt)\, \phi_{bt}(y - x - at), \; b = \sigma^2.$$

The basic iterative formula of order $N/2$

$$p_t^N = p_t^0 + p_t^{N-1} \circledast \Phi, \;\; p_t^0 = \Upsilon_t^{(0)}(a,b), \;\; \Phi = \Phi_a + \Phi_b.$$

where Φ is an auxiliary kernel, $\circledast$ is the space-time convolution.

$$\Phi_b := \frac{1}{2}(b' - b)\Upsilon_t^{(2)}(a,b) =$$
$$\frac{1}{2}\sum_{k=1}^{N} \frac{(-1)^k}{k!}(y-x)^k \Upsilon_t^{(2)}(a,b)\partial_y^k b + R\left(N/2 + \gamma/2 - 1\right),$$

$$\Phi_a := (a' - a)\Upsilon_t^{(1)}(a,b) =$$
$$\sum_{k=1}^{N-1} \frac{(-1)^k}{k!}(y-x)^k \Upsilon_t^{(1)}(a,b)\partial_y^k a + R\left(N/2 + \gamma/2 - 1\right).$$

Origin term decomposition

$$p_t^0 := \Upsilon_t^{(0)}(a,b) = \Upsilon_t^{(0)}(a',b') +$$

$$\sum_{k=1}^{N} \frac{t^k}{k!2^k} \Upsilon_t^{(2k)}(a',b') \left(\sum_{m=1}^{[N/k]} \frac{(y-x)^m}{m!} \partial_x^m b' \right)^k +$$

$$\sum_{k=1}^{N} \frac{1}{k!2^k} \left(\sum_{m=1}^{[N/k]} \frac{(y-x)^m}{m!} \partial_x^m a' \right)^k \cdot$$

$$\sum_{i=1}^{N} \frac{t^{i+k}}{i!} \Upsilon_t^{(2i+k)}(a',b') \cdot \left(\sum_{j=1}^{[N/k-1]} \frac{(y-x)^j}{j!} \partial_x^j a' \right)^i +$$

$$\sum_{k=1}^{N} \frac{t^k}{k!} \Upsilon_t^{(k)}(a',b') \left(\sum_{m=1}^{[N/k-1]} \frac{(y-x)^m}{m!} \partial_x^m a' \right)^k + R\left(N/2 + \gamma/2\right).$$

Algorithm 1

1. If there is at least one difference in the syntax tree $y - x$, then go to step 2, otherwise, go to step 6.
2. Perform the transformation of the expression represented by an abstract syntax tree into a polynomial using the rules of multiplication of polynomials and using a polynomial theorem for increasing polynomials to an integer degree.
3. Simplify the abstract syntax tree.
4. While in a syntax tree it is possible to select at least one product of the Hermite function Υ on $y - x$, perform its transformation using the formula below.
5. Go to step 1.
6. Simplify the abstract syntax tree.
7. Remove terms with order greater than $N/2$. The order of the term refers to the degree of variable t.
8. End.

$$(y - x)\,\Upsilon_t^{(m)}(a, b) = a' t \Upsilon_t^{(m)}(a, b) + b t \Upsilon_t^{(m+1)}(a, b) + m \Upsilon_t^{(m-1)}(a, b).$$

Convolution programming

$$\int_0^t A_1(x)\, s^{k_1} \Upsilon_s^{(m_1)}(a', b') * A_2(y)\, (t-s)^{k_2}\, \Upsilon_{t-s}^{(m_2)}(a, b)\, ds =$$

$$= A_1(x) A_2(y)\, t^{k_1+k_2+1} \Big(\Upsilon_t^{(m)}(a', b')\, B(k_1+1, k_2+1) +$$

$$\sum_{k=1}^{M1} \frac{t^k}{k!2^k} \sum_{j=1}^{M2} \frac{t^j}{j!} \Upsilon_t^{(m+2k+j)}(a', b')\, (a'-a)^j\, (b'-b)^k \times$$

$$B(k_1+k+j+1, k_2+1) +$$

$$\sum_{k=1}^{M3} \frac{t^k}{k!2^k} \Upsilon_t^{(m+2k)}(a', b')\, (b'-b)^k\, B(k_1+k+1, k_2+1) +$$

$$\sum_{k=1}^{M4} \frac{t^k}{k!}\, \Upsilon_t^{(m+k)}(a', b')\, (a'-a)^k\, B(k_1+k+1, k_2+1) \Big) + R.$$

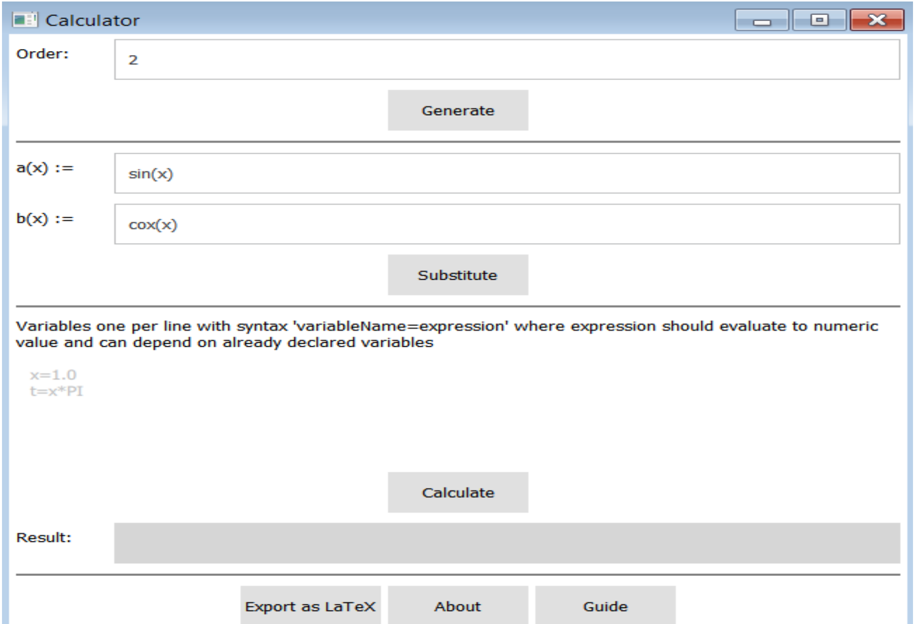
Algorithm 2

1. Let P — set of terms of the first polynomial and F — set of terms of the second polynomial.
2. Find the set of Cartesian product $S = P \times F$.
3. For each element S apply the formula above. Find the sum of all results.
4. Decompose the functions that occur as a result and depend on y in the Taylor series at the point x.
5. Perform a tree simplification that is the result of the transformations.
6. End.

Algorithm 3

1. Using **Algorithm 1** rewrite in terms of Hermite functions a zero term p_t^0 and auxiliary kernel Φ.
2. $i := 1$.
3. Using **Algorithm 2** to find polynomial $p_t'^i = p_t^{i-1} \circledast \Phi$.
4. Using **Algorithm 1** rewrite $p_t'^i$ in terms of Hermite functions.
5. Find $p_t^i = p_t^0 + p_t'^i$.
6. Do simplification p_t^i.
7. Remove from polynomial p_t^i terms with an order greater than $N/2$.
8. If $i < N$, then go to step 9, otherwise, go to step 10.
9. $i := i + 1$.
10. End.

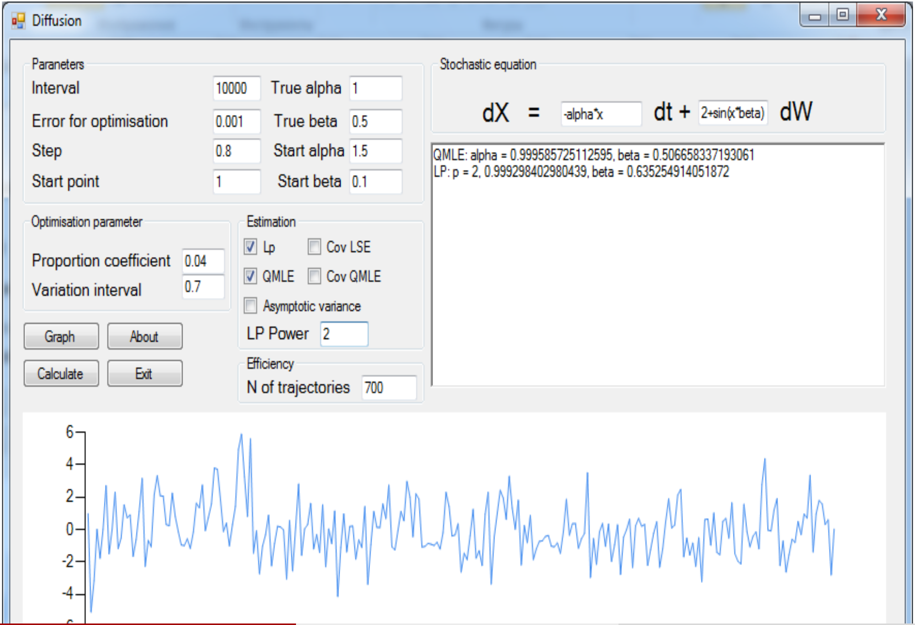
Programm interface



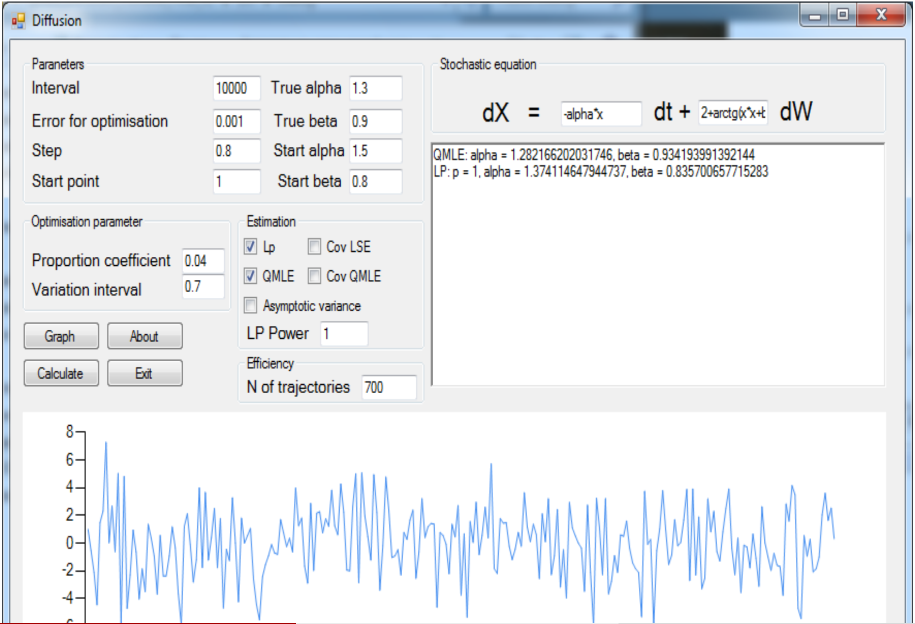
Result of execution

$$
\begin{aligned}
p_t^2 = \Upsilon_t^{(0)}(a,b) + \frac{1}{4} \cdot b \cdot \partial_x(b) \cdot t^2 \cdot \Upsilon_t^{(3)}(a,b) + \\
\left(\frac{1}{2} \cdot b \cdot \partial_x(a) + \frac{1}{8} \cdot b \cdot \partial_x^2(b)\right) \cdot t^2 \cdot \Upsilon_t^{(2)}(a,b) + \\
\left(\frac{1}{4} \cdot a \cdot \partial_x(b) + \frac{3}{2} \cdot (\partial_x(b))^2\right) \cdot t^2 \cdot \Upsilon_t^{(2)}(a,b) + \\
\left(\frac{5}{4} \cdot b \cdot (\partial_x(b))^2 + \frac{1}{12} \cdot \partial_x^2(b) \cdot (b)^2\right) \cdot t^3 \cdot \Upsilon_t^{(4)}(a,b) + \\
\frac{5}{32} \cdot (\partial_x(b))^2 \cdot (b)^2 \cdot t^4 \cdot \Upsilon_t^{(6)}(a,b).
\end{aligned}
$$

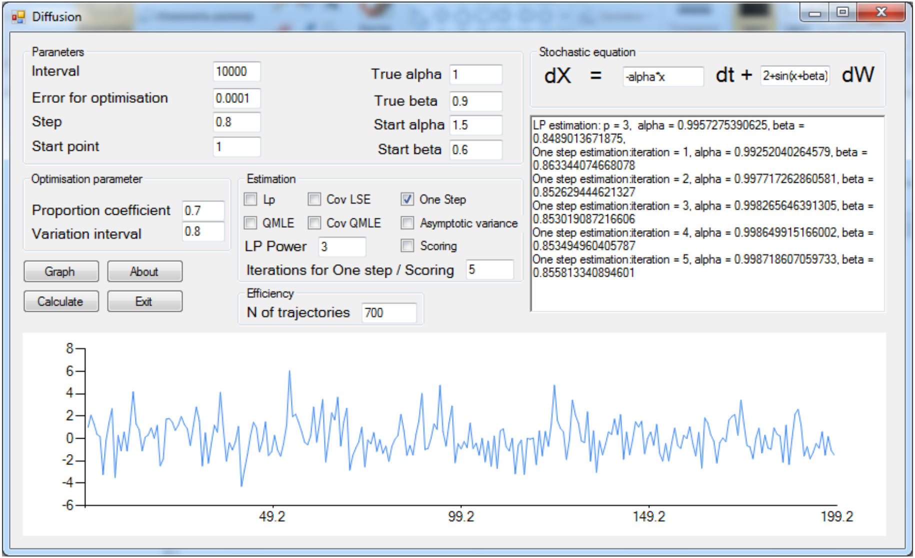
QMLE and LSE


Diffusion
Paraneters
Interval 10000
True alpha 1
Error for optimisation 0.001
True beta 0.5
Step 0.8
Start alpha 1.5
Start point 1
Start beta 0.1
Stochastic equation
dX = -alpha*x dt + 2+sin(x*beta) dW
QMLE: alpha = 0.999585725112595, beta = 0.506658337193061
LP: p = 2, 0.999298402980439, beta = 0.635254914051872
Optimisation parameter
Proportion coefficient 0.04
Variation interval 0.7
Estimation
Lp
Cov LSE
QMLE
Cov QMLE
Asymptotic variance
LP Power 2
Graph
About
Calculate
Exit
Efficiency
N of trajectories 700
6
4
2
0
-2
-4

QMLE and LAE


Diffusion
Paraneters
Interval
10000
True alpha
1.3
Error for optimisation
0.001
True beta
0.9
Step
0.8
Start alpha
1.5
Start point
1
Start beta
0.8
Stochastic equation
dX = -alpha*x dt + 2+arctg(x*x+t dW
QMLE: alpha = 1.282166202031746, beta = 0.934193991392144
LP: p = 1, alpha = 1.374114647944737, beta = 0.835700657715283
Optimisation parameter
Proportion coefficient
0.04
Variation interval
0.7
Estimation
Lp
Cov LSE
QMLE
Cov QMLE
Asymptotic variance
LP Power
1
Graph
About
Calculate
Exit
Efficiency
N of trajectories
700
8
6
4
2
0
-2
-4

One-Step estimator (Newton‘s method)



One-Step estimator (Rao's accumulation method)