TMS165/MSA350 Stochastic Calculus,

Lecture on Applications

In this lecture we demonstrate how statistical methods such as the maximum likelihood method and likelihood ratio estimation can be applied to the Ornstein-Uhlenbeck (OU) process.

1. Elements of diffusion theory

Diffusion processes and SDE

A time homogeneous diffusion process is the solution $X = \{X(t)\}_{t \geq 0}$ to an SDE of the form

$$dX(t) = \mu(X(t))\,dt + \sigma(X(t))\,dB(t) \quad \text{for } t \geq 0, \tag{1}$$

where the drift coefficient $\mu : \mathbb{R} \to \mathbb{R}$ and the diffusion coefficient $\sigma : \mathbb{R} \to \mathbb{R}$ are "sufficiently nice" functions. Here $B = \{B(t)\}_{t \geq 0}$ denotes a Brownian motion as usual.

By definition, the solution X to (1) satisfies

$$X(t) = X(0) + \int_0^t \mu(X(r))\,dr + \int_0^t \sigma(X(r))\,dB(r) \quad \text{for } t \geq 0. \tag{2}$$

Markov property

The solution X to the SDE (1) is a Markov process, which is to say that

$$\mathbf{P}\big\{X(t) \in \cdot \,\big|\, \mathcal{F}_s^X\big\} = \mathbf{P}\{X(t) \in \cdot \,|\, X(s)\} \quad \text{for } 0 \leq s \leq t.$$

Here $\mathcal{F}_s^X = \sigma\{X(r) : r \leq s\}$ for $s \geq 0$ is the filtration generated by the process X itself.

While a detailed rigorous proof of the Markov property is extremly complicated, the Markov property is easy to understand from a more heursitic point of view: Using the representation (2) for both $X(t)$ and $X(s)$ we get

$$\begin{aligned} X(t) &= X(s) + \int_s^t \mu(X(r))\,dr + \int_s^t \sigma(X(r))\,dB(r) \\ &= X(s) + \lim \sum_{i=1}^n \mu(X(t_{i-1}))\,(t_i - t_{i-1}) + \lim \sum_{i=1}^n \sigma(X(t_{i-1}))\,(B(t_i) - B(t_{i-1})), \end{aligned}$$

where $s = t_0 < t_1 < \ldots < t_n = t$ is a partion of the interval $[s, t]$ that becomes finer and finer in the limit. From this we see the that the only thing from the past $\mathcal{F}_s^X$ that affects the future value $X(t)$ is the value $X(s) = X(t_0)$ of X at time $s = t_0$.

Transition densities

The transition density function

$$p(t,x,y) = \frac{d}{dy}\,\mathbf{P}\{X(t{+}s) \leq y\,|\,X(s) = x\} \quad \text{for } t > 0$$

of the diffusion process X given by (1) or (2) satisfies the Kolmogorov backward PDE

$$\frac{\partial}{\partial t}\,p(t,x,y) = \frac{\sigma(x)^2}{2}\,\frac{\partial^2}{\partial x^2}\,p(t,x,y) + \mu(x)\,\frac{\partial}{\partial x}\,p(t,x,y).$$

Conversely, under general conditions, a solution to this PDE is the transition density function of the diffusion process given by (1) or (2) if it is a probability density function as a function of y for each choice of x and $t>0$, that is,

$$p(t,x,y) \geq 0 \quad \text{and} \quad \int_{\mathbb{R}} p(t,x,y)\,dy = 1,$$

and in addition satisfies $p(t,x,y) \to 0$ as $t \downarrow 0$ for $x \neq y$.

In general it is not an easy task or even possible to find an explicit expression for the transition density function. Argubly, the most common systematic way to try to solve the Kolmogorov backward PDE is to consider the Laplace transform

$$\hat{p}(\lambda,x,y) = \int_0^\infty \mathrm{e}^{-\lambda t}\,p(t,x,y)\,dt \quad \text{for } \lambda > 0,$$

which must satisfy the ODE

$$-\lambda\,\hat{p}(\lambda,x,y) = \frac{\sigma(x)^2}{2}\,\frac{\partial^2}{\partial x^2}\,\hat{p}(\lambda,x,y) + \mu(x)\,\frac{\partial}{\partial x}\,\hat{p}(\lambda,x,y)$$

(for $x \neq y$), as

$$\int_0^\infty \mathrm{e}^{-\lambda t}\,\frac{\partial}{\partial t}\,p(t,x,y)\,dt = -\lambda\int_0^\infty \mathrm{e}^{-\lambda t}\,p(t,x,y)\,dt$$

(for $x \neq y$) when $p(0,x,y) = 0$. The conditions that p is a density function translates to

$$\int_{\mathbb{R}} \hat{p}(\lambda,x,y)\,dy = \int_{\mathbb{R}}\int_0^\infty \mathrm{e}^{-\lambda t}\,p(t,x,y)\,dtdy = \int_0^\infty \mathrm{e}^{-\lambda t}\int_{\mathbb{R}} p(t,x,y)\,dydt = \int_0^\infty \mathrm{e}^{-\lambda t}\,dt = \frac{1}{\lambda}.$$

As the ODE for $\hat{p}(\lambda,x,y)$ usually has a unique solution that integrates to $1/\lambda$, in the above fashion, this determines the Laplace transform $\hat{p}(\lambda,x,y)$ of $p(t,x,y)$, after which $p(t,x,y)$ is found by inverse Laplace tranformation.

Although the above indicated way to find transition densities for diffusions does indeed work for many important equations, the details of the solution are in general too difficult to be attempted on undergraduate level, and it is usually more rewarding to search the literature (web) for solutions than to try to derive them oneself.

Stationary distribution

A stationary density function for X is a probability density function π that satisfies

$$\pi(y) = \int_{\mathbb{R}} \frac{d}{dy}\,\mathbf{P}\{X(t+s) \le y \,|\, X(s) = x\}\,\pi(x)\,dx = \int_{\mathbb{R}} p(t,x,y)\,\pi(x)\,dx \quad \text{for } t > 0.$$

By the Chapman-Kolmogorov equation below Theorem 5.6 in Klebaner's book, this means that if the process has the stationary density function (distribution) at a certain time, then it has the stationary density function (distribution) at all later times. It can be shown (see Task 2 of Home exercise session 5) that if X is started according to the stationary distribution at time $t = 0$, then X is a stationary process. Also, regardless of how X is started, X will converge (in a way we select not to specify) to a stationary process with the stationary marginal distribution as $t \to \infty$.

As the transition density satisfies the Kolmogorov forward PDE

$$\frac{\partial}{\partial t}\,p(t,x,y) = \frac{\partial^2}{\partial y^2}\bigg(\frac{\sigma(y)^2}{2}\,p(t,x,y)\bigg) - \frac{\partial}{\partial y}\Big(\mu(y)\,p(t,x,y)\Big),$$

we see from the above integral equation that

$$\begin{aligned}
-\bigg(\frac{\partial^2}{\partial y^2}\,\frac{\sigma(y)^2}{2} - \frac{\partial}{\partial y}\,\mu(y)\bigg)\pi(y) &= \bigg(\frac{\partial}{\partial t} - \frac{\partial^2}{\partial y^2}\,\frac{\sigma(y)^2}{2} + \frac{\partial}{\partial y}\,\mu(y)\bigg)\pi(y) \\
&= \int_{\mathbb{R}} \bigg(\frac{\partial}{\partial t} - \frac{\partial^2}{\partial y^2}\,\frac{\sigma(y)^2}{2} + \frac{\partial}{\partial y}\,\mu(y)\bigg)p(t,x,y)\,\pi(x)\,dx \\
&= 0.
\end{aligned}$$

If it exists, then the stationary density function is given by

$$\pi(y) = \frac{C}{\sigma(y)^2}\,\exp\bigg\{\int_{y_0}^{y} \frac{2\,\mu(z)}{\sigma(z)^2}\,dz\bigg\}.$$

Here $C > 0$ is a normalizing constant selected to make π a probability density function, that is, $\int_{\mathbb{R}} \pi(y)\,dy = 1$, while $y_0 \in \mathbb{R}$ is any constant. The existence of the stationary density function is the same thing as that this normalizing procedure is possible to carry out. Note that it is easy to see that π satisfies the ODE (cf. above)

$$\bigg(\frac{\partial^2}{\partial y^2}\,\frac{\sigma(y)^2}{2} - \frac{\partial}{\partial y}\,\mu(y)\bigg)\pi(y).$$

Finite dimensional distributions

The joint probability density function of $(X(t_0), X(t_1), \ldots, X(t_n))$ for $0 = t_0 < t_1 < \ldots < t_n$ is given by

$$f_{X(t_0),X(t_1),\ldots,X(t_n)}(x_0, x_1, \ldots, x_n) = \pi(x_0) \prod_{i=1}^{n} p(t_i - t_{i-1}, x_{i-1}, x_i)$$

when the process is started with the stationary density function π (provided that it exists) at time 0, while

$$f_{X(t_1),\ldots,X(t_n)}(x_1,\ldots,x_n) = \prod_{i=1}^{n} p(t_i - t_{i-1}, x_{i-1}, x_i)$$

when the process is started at a fixed value $X(0) = x_0$.

Euler method

We may simulate a weak solution to the SDE at a time grid $0 = t_0 < t_1 < \ldots < t_n = T$ by means of the Euler method, as

$$X(t_i) \approx X(t_{i-1}) + \mu(X(t_{i-1}))\,(t_i - t_{i-1}) + \sigma(X(t_{i-1}))\,\xi_i \quad \text{for } i = 1,\ldots,n,$$

where $\{\xi_i\}_{i=1}^n$ are independent random variables such that ξ_i is zero-mean normal distributed with standarddeviation $\sqrt{t_i - t_{i-1}}$, and where $X(0)$ is a random variable that is indepedent of $\{\xi_i\}_{i=1}^n$ and has the stationary distribution if X is started according to the stationary distribution, while $X(0) = x_0$ if X is started at a fixed value x_0.

Likelihood ratios

If the diffusion process X satisfies the equation

$$dX(t) = \mu_1(X(t))\,dt + \sigma(X(t))\,dB(t)$$

for a $\mathbf{P}_1$-Brownian motion B and

$$dX(t) = \mu_2(X(t))\,dt + \sigma(X(t))\,dW(t)$$

for a $\mathbf{P}_2$-Brownian motion W, where $\mathbf{P}_1$ and $\mathbf{P}_2$ are two different probability measures, then the likelihood ratio between $\mathbf{P}_2$ and $\mathbf{P}_1$ based on an observation $\{X(t)\}_{t\in[0,T]}$ of the process X in the time interval $[0,T]$ is given by

$$\frac{d\mathbf{P}_2}{d\mathbf{P}_1} = \exp\left\{ \int_0^T \frac{\mu_2(X(t)) - \mu_1(X(t))}{\sigma(X(t))^2}\,dX(t) - \frac{1}{2}\int_0^T \frac{\mu_2(X(t))^2 - \mu_1(X(t))^2}{\sigma(X(t))^2}\,dt \right\}.$$

This likelihood ratio can be used to judge which is the most likely of the above two SDE-models for X based on the observations: If $d\mathbf{P}_2/d\mathbf{P}_1$ is bigger than 1 then the model with the drift μ_2 is the most likely, while a value of $d\mathbf{P}_2/d\mathbf{P}_1$ that is smaller than 1 indicates that the drift μ_1 is the most likely.

The likelihood ratio can also be used to estimate parameters of a parametric SDE, as we will see later.

2. The OU process

An OU process is the solution X to the SDE

$$dX(t) = -\mu\, X(t)\, dt + \sigma\, dB(t),$$

where $\mu > 0$ and $\sigma > 0$ are parameters. In other words, the drift is $\mu(x) = -\mu\, x$ and the difffusion coefficient $\sigma(x) = \sigma$.

The OU process has stationary distribution

$$\pi(y) = \frac{1}{\sigma^2}\exp\biggl\{-\int_0^y \frac{2\,\mu\, z}{\sigma^2}\, dz\biggr\}\bigg/\biggl(\int_{\mathbb{R}} \frac{1}{\sigma^2}\exp\biggl\{-\int_0^y \frac{2\,\mu\, z}{\sigma^2}\, dz\biggr\}\, dy\biggr) = \frac{\sqrt{\mu}}{\sqrt{\pi}\,\sigma}\exp\biggl\{-\frac{\mu\, y^2}{\sigma^2}\biggr\},$$

that is, a zero-mean normal distribution with standarddeviation $\sigma/\sqrt{2\,\mu}$.

As we have shown in Exercise Session 5, the OU process has transition density

$$p(t,x,y) = \frac{\sqrt{\mu}}{\sqrt{\pi}\,\sigma\sqrt{1-\mathrm{e}^{-2\mu t}}}\exp\biggl\{-\frac{(y-\mathrm{e}^{-\mu t}x)^2}{\sigma^2\,(1-\mathrm{e}^{-2\mu t})/\mu}\biggr\},$$

that is, a normal distribution with mean $\mathrm{e}^{-\mu t}x$ and standarddeviation $\sigma\sqrt{1-\mathrm{e}^{-2\mu t}}/\sqrt{2\,\mu}$. It is also a routine matter to differentiate to check that this function p satisfies the Kolmogorov backward PDE.

If X is an OU process

$$dX(t) = -\mu_1\, X(t)\, dt + \sigma\, dB(t)$$

for a $\mathbf{P}_{\mu_1}$-Brownian motion B, and an OU process

$$dX(t) = -\mu_2\, X(t)\, dt + \sigma\, dW(t)$$

for a $\mathbf{P}_{\mu_2}$-Brownian motion W, then the corresponding likelihood ratio is given by

$$\frac{d\mathbf{P}_{\mu_2}}{d\mathbf{P}_{\mu_1}} = \exp\biggl\{-\frac{\mu_2-\mu_1}{\sigma^2}\int_0^T X(t)\, dX(t) - \frac{\mu_2^2-\mu_1^2}{2\,\sigma^2}\int_0^T X(t)^2\, dt\biggr\}.$$

In particular, we can find which is the most likely of the models

$$dX(t) = dB(t)$$

and

$$dX(t) = -X(t)\, dt + dW(t)$$

by computing the likelihood ratio for $\mu_1 = 0$, $\mu_2 = 1$ and $\sigma = 1$

$$\frac{d\mathbf{P}_1}{d\mathbf{P}_0} = \exp\biggl\{-\int_0^T X(t)\, dX(t) - \frac{1}{2}\int_0^T X(t)^2\, dt\biggr\},$$

and then check whether $d\mathbf{P}_1/d\mathbf{P}_0 > 1$, indicating that $\mu = 1$ is the most appropriate model, or $d\mathbf{P}_1/d\mathbf{P}_0 < 1$, indicating that $\mu = 0$ is the most appropriate model.

We can also estimate the parameter μ for the equation

$$dX(t) = -\mu\, X(t)\, dt + dW(t)$$

by means of maximizing the likelihood

$$\frac{d\mathbf{P}_\mu}{d\mathbf{P}_0} = \exp\left\{-\mu \int_0^T X(t)\, dX(t) - \frac{\mu^2}{2} \int_0^T X(t)^2\ dt\right\},$$

which by differentiation gives the estimate

$$\mu = -\int_0^T X(t)\, dX(t) \Big/ \left(\int_0^T X(t)^2\ dt\right).$$

3. Application to the OU process

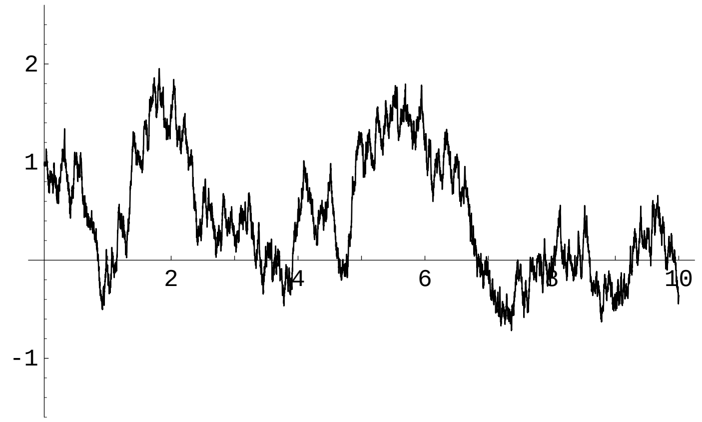
We used the Euler method to simulate an OU process $\{X(t)\}_{t\in[0,10]}$ started according to the stationary distribution, and an OU process $\{Y(t)\}_{t\in[0,10]}$ started at zero. In both cases the drift was $\mu(x) = -\mu_0 x$ and the diffusion coefficient $\sigma(x) = \sigma_0$, where $\mu_0 = \sigma_0 = 1$.

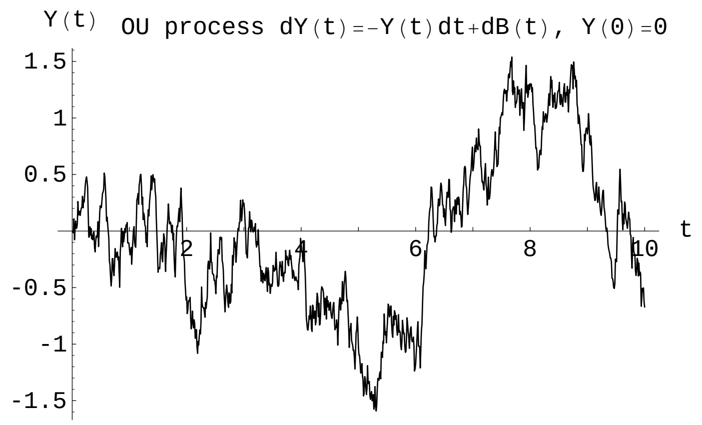
We use distance $\frac{1}{100}$ between the time points of the simulation grid, so that $0 = t_0 < t_1 < \ldots < t_{1000} = 10$, where $t_i - t_{i-1} = \frac{1}{100}$ for $i = 1, \ldots, 1000$.

The simulations were carried out by means of the following Mathematica programs.

```
<<Statistics‘ContinuousDistributions‘;
dt=1/100; T=10; {mu0,sigma0}={1,1};
For[i=2; X={Random[NormalDistribution[0,sigma0/Sqrt[2*mu0]]]},
    i<=T/dt, i++, AppendTo[X, X[[i-1]] - mu0*X[[i-1]]*dt
    + Random[NormalDistribution[0,sigma0*Sqrt[dt]]]]]
For[i=2; Y=0, i<=T/dt, i++, AppendTo[Y, Y[[i-1]] - mu0*Y[[i-1]]*dt
    + Random[NormalDistribution[0,sigma0*Sqrt[dt]]]]]
```

The results of the simulations are depicted in the following two figures





The joint density functions of $(X(t_0),\ldots,X(t_{1000}))$ and $(Y(t_0),\ldots,Y(t_{1000}))$ are given by

$$\begin{aligned}
&f_{X(t_0),\ldots,X(t_{1000})}(x_0,\ldots,x_{1000}) \\
&\qquad = \pi(x_0)\prod_{i=1}^{1000} p(t_i - t_{i-1}, x_{i-1}, x_i) \\
&\qquad = \frac{\sqrt{\mu}}{\sqrt{\pi}\,\sigma}\exp\left\{-\frac{\mu x_0^2}{\sigma^2}\right\}\prod_{i=1}^{1000}\frac{\sqrt{\mu}}{\sqrt{\pi}\,\sigma\sqrt{1-\mathrm{e}^{-2\mu/100}}}\exp\left\{-\frac{(x_i-\mathrm{e}^{-\mu/100}x_{i-1})^2}{\sigma^2\,(1-\mathrm{e}^{-2\mu/100})/\mu}\right\}
\end{aligned}$$

and

$$\begin{aligned}
f_{Y(t_1),\ldots,Y(t_{1000})}(y_1,\ldots,y_{1000}) &= \prod_{i=1}^{1000} p(t_i - t_{i-1}, y_{i-1}, y_i) \\
&= \prod_{i=1}^{1000}\frac{\sqrt{\mu}}{\sqrt{\pi}\,\sigma\sqrt{1-\mathrm{e}^{-2\mu/100}}}\exp\left\{-\frac{(y_i-\mathrm{e}^{-\mu/100}y_{i-1})^2}{\sigma^2\,(1-\mathrm{e}^{-2\mu/100})/\mu}\right\},
\end{aligned}$$

respectively, where $y_0 = 0$. Hence we may use the maximum likelihood method to estimate μ and σ from our simulated data (pretending that they are unknown), by means of maximizing the likelihood

$$\frac{\sqrt{\mu}}{\sqrt{\pi}\,\sigma}\exp\biggl\{-\frac{\mu\,X(t_0)^2}{\sigma^2}\biggr\}\prod_{i=1}^{1000}\frac{\sqrt{\mu}}{\sqrt{\pi}\,\sigma\sqrt{1-\mathrm{e}^{-2\mu/100}}}\exp\biggl\{-\frac{(X(t_i)-\mathrm{e}^{-\mu/100}X(t_{i-1}))^2}{\sigma^2\,(1-\mathrm{e}^{-2\mu/100})/\mu}\biggr\}$$

and

$$\prod_{i=1}^{1000}\frac{\sqrt{\mu}}{\sqrt{\pi}\,\sigma\sqrt{1-\mathrm{e}^{-2\mu/100}}}\exp\biggl\{-\frac{(Y(t_i)-\mathrm{e}^{-\mu/100}Y(t_{i-1}))^2}{\sigma^2\,(1-\mathrm{e}^{-2\mu/100})/\mu}\biggr\},$$

respectively. These maximum likelihood estimates were carried out by means of the following Mathematcia code (with the densities logged to not get numerical underflows).

```
fOUStationary[mu_,sigma_,x_]
     := Sqrt[mu]*Exp[-mu*x^2/sigma^2]/(Sqrt[Pi]*sigma);
pOU[mu_,sigma_,x_,y_,t_]
     := Exp[-(y-x*Exp[-mu*t])^2/(2*sigma^2*(1-Exp[-2*mu*t])/(2*mu))]
         /(Sqrt[2*Pi]*sigma*Sqrt[1-Exp[-2*mu*t]]/Sqrt[2*mu]);
MLStationary[mu_,sigma_,dt_,Data_]
     := Log[fOUStationary[mu,sigma,Data[[1]]]]
         + Sum[Log[pOU[mu,sigma,Data[[i-1]],Data[[i]],dt]],
         {i,2,Length[Data]}]
MLNonStationary[mu_,sigma_,dt_,Data_]
     := Sum[Log[pOU[mu,sigma,Data[[i-1]],Data[[i]],dt]],
         {i,2,Length[Data]}]
NMaximize[MLStationary[mu,sigma,dt,X],mu>0,sigma>0, mu,sigma]
Out[]:= {890.45, {mu -> 1.01156, sigma -> 0.996303}}
NMaximize[MLNonStationary[mu,sigma,dt,Y],mu>0,sigma>0, mu,sigma]
Out[]:= {886.86, {mu -> 0.997098, sigma -> 1.00088}}
```

Note how well this fits with the correct values $\mu=1$ and $\sigma=1$ for the parameters.

We calculate the likelihood ratios

$$\frac{d\mathbf{P}_1}{d\mathbf{P}_0}(X)=\exp\biggl\{-\int_0^{10}X(t)\,dX(t)-\frac{1}{2}\int_0^{10}X(t)^2\;dt\biggr\}$$

and

$$\frac{d\mathbf{P}_1}{d\mathbf{P}_0}(Y)=\exp\biggl\{-\int_0^{10}Y(t)\,dY(t)-\frac{1}{2}\int_0^{10}Y(t)^2\;dt\biggr\},$$

for our simulated processes, in order to fine whether

$$dX(t) = dB(t) \quad \text{or} \quad dX(t) = -X(t)\,dt + dW(t)$$

and

$$dY(t) = dB(t) \quad \text{or} \quad dY(t) = -Y(t)\,dt + dW(t),$$

respectively, are the most likely models for the data. As both ratios where significantly larger than 1 (see the enclosed Mathematica code), the model with $\mu = 1$ was the most likely for both data sets.

```
OURatioTest[Data_]
    := Exp[Sum[-Data[[i-1]]*(Data[[i]]-Data[[i-1]]), {i,2,
        Length[Data]}] - Sum[Data[[i]]^2*dt, {i,1,Length[Data]}]/2];
{OURatioTest[X], OURatioTest[Y]}
Out[]:= {10.1342, 10.5346}
```

We may estimate the parameter μ by means of maximizing the likelihood ratios $(d\mathbf{P}_\mu/d\mathbf{P}_0)(X)$ and $(d\mathbf{P}_\mu/d\mathbf{P}_0)(Y)$, respectively, which gives the μ estimates

$$\mu = -\int_0^{10} X(t)\,dX(t) \bigg/ \left(\int_0^{10} X(t)^2\,dt \right) \quad \text{and} \quad \mu = -\int_0^{10} Y(t)\,dY(t) \bigg/ \left(\int_0^{10} Y(t)^2\,dt \right),$$

respectively. Both results were very close to the correct $\mu = 1$, as the following Mathematica code illustrates:

```
OURatioEst[Data_]
    := -Sum[Data[[i-1]]*(Data[[i]]-Data[[i-1]]), {i,2,Length[Data]}]
        /Sum[Data[[i]]^2*dt, {i,1,Length[Data]}];
{OURatioEst[X], OURatioEst[Y]}
{0.954688, 0.991207}
```