Lecture on Applications Monday 10 December

In this lecture we give a detailed demonstration of SDE and diffusion theory as well as statistical likelihood methodology to the Ornstein-Uhlenbeck (OU) process.

1. Elements of diffusion theory

Diffusion processes

Given "nice" drift and volatility function $\mu : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$, respectively, a time homogeneous diffusion process is the solution $X(t)$ to an SDE of the form

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

Transition densities

The transition density function

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

of the diffusion process 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 if it is a density function for each 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 it satisfies $p(t,x,y) \to 0$ as $t \downarrow 0$ for $x \neq y$.

In general it is not easy to find an explicit expression for the transition density function. The most common way is to solve the PDE by means of Laplace transformation

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

The Laplace transformed density $\hat{p}(\lambda,x,y)$ 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 method to find transition densities works for many important equations, the details are often too difficult to attempt on undergraduate level, and it is more rewarding to search the literature (web) for solutions than to try own derivations.

Stationary distribution

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

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

This means that if the process has the stationary distribution at a ceratin time, then it also has the stationary distribution at all later times. This implies (by some further considerations), that if the process is started with the stationary distribution, then it is a stationary process. Also, if the process is started at a fixed value, then it will converge (in a ceratin sense) to a stationary process with the stationary marginal distribution.

As the transition density satisfies the Kolmogorov forward PDE

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

we see from the above integral equation that

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

If it exists, the stationary density is given by

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

where $y_0 \in \mathbb{R}$ is any constant and $C>0$ is a normalizing constant selected to make π a density, that is, $\int_{\mathbb{R}}\pi(y)\,dy = 1$. The stationary density exists when this normalization can be carried out. Note that it is easy to see that this π satisfies the ODE (cf. above)

$$\Big(\frac{\partial^2}{\partial y^2}\frac{\sigma(y)^2}{2} + \frac{\partial}{\partial y}\mu(y)\Big)\pi(y).$$

Finite dimensional distributions

The joint density function of $(X(t_1),\dots,X(t_n))$, $0<t_1<\dots<t_n$, is given by

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

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

$$f_{X(t_1),\dots,X(t_n)}(x_1,\dots,x_n)=\pi(x_1)\prod_{i=2}^{n} p(t_i-t_{i-1},x_{i-1},x_i)$$

when the process is started with the stationary distribution (provided that it exists).

Euler method

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

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

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

For a simulated approximate strong solution based on a given Brownian motion $B(t)$ we use $B(t_i)-B(t_{i-1})$ instead of $\sqrt{t_i-t_{i-1}}\,\xi_i$ in the above algorithm.

Likelihood ratios

If for two different probability measures $\mathbf{P}_1$ and $\mathbf{P}_2$ the diffusion process $X(t)$ satisfies

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

for a $\mathbf{P}_1$-Brownian motion $B(t)$, and

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

for a $\mathbf{P}_2$-Brownian motion $W(t)$, then the likelihood ratio between $\mathbf{P}_2$ and $\mathbf{P}_1$, based on the process in the time interval $[0,T]$ is given by

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

Given a data set $\{X(t)\}_{t\in[0,T]}$ the likelihood ratio can be used to judge which is the most likely (best) of the above two SDE models for $X(t)$: If $d\mathbf{P}_2/d\mathbf{P}_1$ is (significantly) bigger than 1, then the model with the drift μ_2 is the most likely, while a $d\mathbf{P}_2/d\mathbf{P}_1$ (significantly) 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.

2. The OU process

Given parameters $\mu, \sigma > 0$, an OU process is the solution $X(t)$ to the Langevin SDE

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

In other words, the drift is $\mu(x) = -\mu\, x$ and the volatility $\sigma(x) = \sigma$.

By the above formula for stationary densities, we see that for the OU process

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

that is, the stationary density is zero-mean normal with variance $\sigma^2/(2\,\mu)$.

The OU process has transition density

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

that is, a normal distribution with mean $\mathrm{e}^{-\mu t}x$ and variance $\sigma^2(1-\mathrm{e}^{-2\mu t})/(2\mu)$. Note that, although one might wonder how this formula is derived, it is straightforward to differentiate in order to check that this p satisfies the backward PDE.

A strong solution to the Langevin SDE is given by

$$X(t) = \mathrm{e}^{-\mu t}\bigg(X(0) + \int_0^t \mathrm{e}^{s}\,\sigma\, dB(s)\bigg) \quad \text{for } t \geq 0:$$

This can be seen by mean of direct calculations checking that this $X(t)$ really solves the SDE. Alternatively, we can deduce this fact from the theory for linear SDE's.

If $X(t)$ is an OU process

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

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

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

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

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

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 $\sigma = 1$

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

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 $\mu_0 = \sigma_0 = 1$.

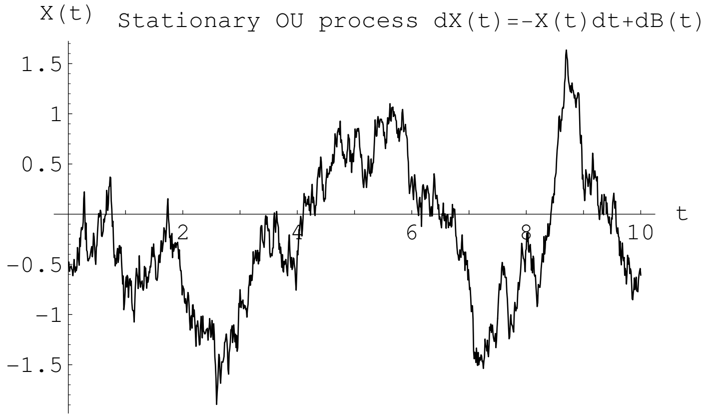
We use distance $\frac{1}{100}$ between the time points of the simulation grid, so that $0 = t_0 < t_1 < \ldots < t_{1000} = 10$, where $t_i - t_{i-1} = \frac{1}{100}$ for $i = 1, \ldots, 1000$.

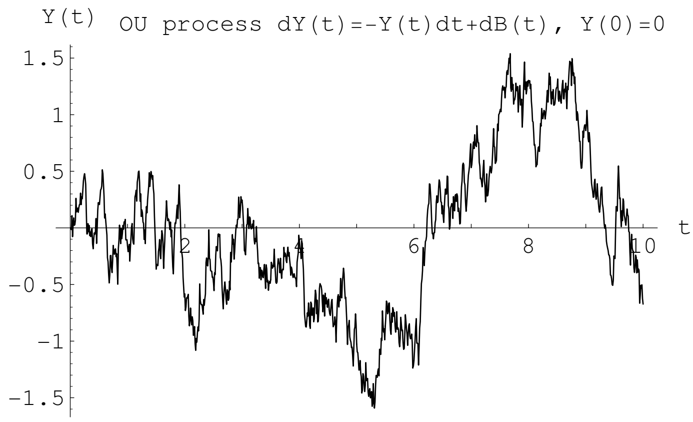
The simulations were carried out by means of the following Mathematica programs.

```
<<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

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

and

$$
\begin{aligned}
f_{Y(t_0),\ldots,Y(t_{1000})}(y_0,\ldots,y_{1000}) &= \prod_{i=1}^{1000} p(t_i-t_{i-1},y_{i-1},y_i)\\
&= \prod_{i=1}^{1000} \frac{\sqrt{\mu}}{\sqrt{\pi}\,\sigma\sqrt{1-\mathrm{e}^{-2\mu/100}}}\exp\left\{-\frac{(y_i-\mathrm{e}^{-\mu/100}y_{i-1})^2}{\sigma^2\,(1-\mathrm{e}^{-2\mu/100})/\mu}\right\},
\end{aligned}
$$

respectively. 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

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

and

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

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

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

and

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

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

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

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}
```