Statistical Inference for Diffusion Processes

Radu Herbei

The Ohio State University
The Mathematical Biosciences Institute

July 7, 2015
Quick review

- $(\Omega, \mathcal{F}, \mathbb{P}), \{\mathbb{W}_t, t \in [0, T]\}$ is a SBM, $\mathcal{F}_t$ - standard filtration
- If a stochastic process $(h_t), t \in [0, T]$ is carefully chosen, one can define the Itô integral
  \[
  \int_0^T h_t \, d\mathbb{W}_t
  \]
- A homogeneous diffusion process $(X_t)$ is characterized via
  \[
  dX_t = S(X_t)dt + \sigma(X_t)d\mathbb{W}_t \quad X(0) = X_0, \ t \in [0, T]
  \]
- Weak solutions, existence, uniqueness, Girsanov formula.
What’s next

- Numerical methods for SDEs
- Ergodicity, LLN, CLTs
- Statistical inference for SDEs (high frequency data)
- Statistical inference for SDEs (discrete data)
Numerical methods for SDEs

A “discretization” of a stochastic process $X^T = \{X_t, t \in [0, T]\}$ is a process $X^T_\delta$, which “approximates” $X^T$ in some way.

The discretization $X_\delta$ has strong order of convergence $\gamma$ if for any $T > 0$,

$$E(|X_\delta(T) - X(T)|) \leq c\delta^\gamma \quad \text{for all} \quad \delta < \delta_0$$

The discretization $X^T_\delta$ has weak order of convergence $\beta$ if for any function $g$ which is $2(\beta + 1)$ continuously differentiable, it is true that

$$|E(g(X_\delta(T))) - E(g(X(T)))| \leq c\delta^\beta \quad \text{for all} \quad \delta < \delta_0$$
Euler scheme

\[ dX_t = S(t, X_t)dt + \sigma(t, X_t)dW_t \]

Given a collection of time points \(0 = t_0 < t_1 \cdots < t_n = T\), the Euler discretization is the process \(X^E(\cdot)\) such that

\[ X^E(t_{i+1}) = X^E(t_i) + S(t_i, X^E(t_i))(t_{i+1} - t_i) + \sigma(t_i, X^E(t_i))(W(t_{i+1}) - W(t_i)) \]

For \(t \in [t_i, t_{i+1})\), the process \(X^E(t)\) can be defined in any way – typically by linear interpolation.

The Euler scheme has strong order of convergence \(\gamma = \frac{1}{2}\) and weak order \(\beta = 1\).
Euler scheme

- Most popular approach to approximate a diffusion process
- Vast majority of research using SDE models actually use the Euler approximation for the computing implementation
- Very easy to implement; very fast to simulate
  - Simulate \( Z \sim \mathcal{N}(0, 1) \)
  - Given \( X^E(t_i) \), set

\[
X^E(t_{i+1}) = X^E(t_i) + S(t_i, X^E(t_i))(t_{i+1} - t_i) + \sqrt{t_{i+1} - t_i}\sigma(t_i, X^E(t_i))Z
\]

- With a little “trick”, can achieve weak order of convergence \( \beta = 1 \).
Milstein scheme

\[ X^M(t_{i+1}) = \]
\[ = X^M(t_i) + S(t_i, X^M(t_i))(t_{i+1} - t_i) + \sigma(t_i, X^M(t_i))(\mathbb{W}(t_{i+1}) - \mathbb{W}(t_i)) \]
\[ + \frac{1}{2} \sigma(t_i, X^M(t_i))\sigma_x(t_i, X^M(t_i)) \left( (\mathbb{W}(t_{i+1}) - \mathbb{W}(t_i))^2 - (t_{i+1} - t_i) \right) \]

Milstein scheme has strong order of convergence \( \gamma = 1 \) and weak order of convergence \( \beta = 1 \).

Note that for the OU process, \( \sigma_x \equiv 0 \) thus, the Euler and Milstein schemes are identical.
Geometric BM

\[ dX_t = \theta_1 X_t dt + \theta_2 X_t d\mathbb{W}_t \]

Euler scheme:

\[ X^E_{i+1} = X^E_i (1 + \theta_1 \Delta) + \theta_2 X^E_i \sqrt{\Delta Z} = X^E_i (1 + \theta_1 \Delta + \theta_2 \sqrt{\Delta Z}) \]

Milstein scheme

\[ X^M_{i+1} = X^M_i + \theta_1 X^M_i \Delta + \theta_2 X^M_i \sqrt{\Delta Z} + \frac{1}{2} \theta_2^2 X^M_i (\Delta Z^2 - \Delta) \]
\[ = X^M_i \left( 1 + \left( \theta_1 + \frac{1}{2} \theta_2^2 (Z^2 - 1) \right) \Delta + \theta_2 \sqrt{\Delta Z} \right) \]

Exact solution

\[ X_{t+\Delta} = X_t \exp \left\{ \left( \theta_1 - \frac{\theta_2^2}{2} \right) \Delta + \theta_2 \sqrt{\Delta Z} \right\} \]
\[ = X_t \left( 1 + \left( \theta_1 - \frac{\theta_2^2}{2} \right) \Delta + \theta_2 \sqrt{\Delta Z} + \frac{1}{2} \theta_2^2 \Delta Z^2 + O(\Delta) \right) \]
Connection between Euler and Milstein schemes

\[ dX_t = S(t, X_t)dt + \sigma(t, X_t)d\mathbb{W}_t \]

Consider the Lamperti transformation \( Y_t = F(X_t) \) where

\[ F(x) = \int_x^\infty \frac{1}{\sigma(t, u)} du \]

Note that

\[ F'(x) = \frac{1}{\sigma(t, x)} \quad F''(x) = -\frac{\sigma_x(t, x)}{\sigma(t, x)^2} \]

Use Itô formula to derive that

\[ dY_t = \left( \frac{S(t, X_t)}{\sigma(t, X_t)} - \frac{1}{2} \sigma_x(t, X_t) \right) dt + d\mathbb{W}_t \]
The Euler scheme for the transformed process gives

\[ \Delta Y = Y_{i+1} - Y_i = \left( \frac{S(t_i, X_i)}{\sigma(t_i, X_i)} - \frac{1}{2} \sigma_x(t_i, X_i) \right) \Delta t + \sqrt{\Delta t} Z \]

Write the Taylor expansion for the inverse transformation \( X = G(Y) \)

\[ G(Y + \Delta Y) = G(Y) + G'(Y) \Delta Y + \frac{1}{2} G''(Y) \Delta Y^2 + \mathcal{O}(\Delta Y^3) \]

where

\[ G'(y) = \frac{1}{F'(G(y))} = \sigma(t, G(y)) \]

\[ G''(y) = G'(y) \sigma_x(t, G(y)) = \sigma(t, G(y)) \sigma_x(t, G(y)) \]

This gives

\[ G(Y_i + \Delta Y) - G(Y_i) = \left( \text{after some algebra} \right) \]

\[ = \left( S - \frac{1}{2} \sigma_x \sigma \right) \Delta + \sigma \sqrt{\Delta} Z + \frac{1}{2} \sigma \sigma_x \Delta Z^2 + \ldots \]

\[ = \text{Milstein for the} \ X \ \text{process} \]
CIR process

\[dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 \sqrt{X_t} d\mathbb{W}_t\]

Milstein scheme (after some re-ordering):

\[X_{i+1} = X_i + \left( (\theta_1 - \theta_2 X_i) - \frac{1}{4} \theta_3^2 \right) \Delta + \theta_3 \sqrt{X_i} \sqrt{\Delta} Z + \frac{1}{4} \theta_3^2 \Delta Z^2\]

Transform using \(Y = F(X) = \sqrt{X}\). Itô formula gives

\[dY_t = \frac{1}{2Y_t} \left( (\theta_1 - \theta_2 Y_t^2) - \frac{1}{4} \theta_3^2 \right) dt + \frac{1}{2} \theta_3 d\mathbb{W}_t\]

for which the Euler scheme is

\[\Delta Y = \frac{1}{2Y_i} \left( (\theta_1 - \theta_2 Y_i^2) - \frac{1}{4} \theta_3^2 \right) \Delta + \frac{1}{2} \theta_3 \sqrt{\Delta} Z\]
CIR process

The inverse transformation is \( x = G(y) = y^2 \). Then

\[
G(Y_i + \Delta Y) - G(Y_i) = (Y_i + \Delta Y)^2 - Y_i^2 = (\Delta Y)^2 + 2Y_i\Delta Y
\]

Replace \( \Delta Y \) with the expression above, ignore high order terms, obtain

\[
\Delta X = G(Y_i + \Delta Y) - G(Y_i)
= \left( \theta_1 - \theta_2 X_i - \frac{1}{4} \theta_3^2 \right) \Delta + \theta_3 \sqrt{X_i} \sqrt{\Delta Z} + \frac{1}{4} \theta_3^2 \Delta Z^2 + ... 
\]

(Milstein scheme for the \( X \) process . . .)
Predictor-corrector method

Tries to correct the fact that $S(t, X_t)$ and $\sigma(t, X_t)$ are not constant on the interval $[t_i, t_{i+1})$.

**Step 1**: predictor:

$$\tilde{X}_{i+1} = X_i + S(t_i, X_i)\Delta t + \sigma(t_i, X_i)\sqrt{\Delta t}Z;$$

**Step 2**: corrector:

$$X_{i+1} = X_i + \left(\alpha \tilde{S}(t_{i+1}, \tilde{X}_{i+1}) + (1-\alpha) \tilde{S}(t_i, X_i)\right)\Delta t$$
$$+ \left(\eta \sigma(t_{i+1}, \tilde{X}_{i+1}) + (1-\eta) \sigma(t_i, X_i)\right) \sqrt{\Delta t}Z$$

where

$$\tilde{S}(t_i, X_i) = S(t_i, X_i) - \eta \sigma(t_i, X_i)\sigma_x(t_i, X_i)$$

This reduces to Euler method when $\alpha = \eta = 0$. 
Second Milstein scheme

\[ X_{i+1} = X_i + \left( S(t_i, X_i) - \frac{1}{2} \sigma(t_i, X_i) \sigma_x(t_i, X_i) \right) \Delta t + \sigma(t_i, X_i) \sqrt{\Delta t} \, Z \]
\[ + \frac{1}{2} \sigma(t_i, X_i) \sigma_x(t_i, X_i) \Delta t \, Z^2 \]
\[ + \Delta t^{3/2} \left( \frac{1}{2} S(t_i, X_i) \sigma_x(t_i, X_i) + \frac{1}{2} S_x(t_i, X_i) \sigma(t_i, X_i) + \frac{1}{4} \sigma(t_i, X_i)^2 \sigma_{xx}(t_i, X_i) \right) \, Z \]
\[ + \Delta t^2 \left( \frac{1}{2} S(t_i, X_i) S_x(t_i, X_i) + \frac{1}{4} S_{xx}(t_i, X_i) \sigma(t_i, X_i)^2 \right) \]

It has weak order of convergence \( \beta = 2 \).
Local linearization methods

Euler method assumes that the drift and diffusion coefficients and constant on small time intervals;
Assume that locally, the drift and diffusion coefficients are linear;

**Ozaki method**

\[
dX_t = b(X_t)dt + \sigma d\mathbb{W}_t
\]

Start with the corresponding deterministic system

\[
\frac{dx_t}{dt} = b(x_t)
\]

which admits the following numerical approximation

\[
x_{t+\Delta t} = x_t + \frac{b(x_t)}{b_x(x_t)} \left( e^{b_x(x_t)\Delta t} - 1 \right)
\]

Assuming that \( b(x) = K_t x \) on the interval \([t, t + \Delta t]\) gives

\[
X_{t+\Delta t} = X_t e^{K_t \Delta t} + \sigma \int_t^{t+\Delta t} e^{K_t(t+\Delta t-u)} d\mathbb{W}_u
\]
The constant $K_t$ is determined from the assumption

$$E(X_{t+\Delta t} \mid X_t) = X_t e^{K_t \Delta t} = X_t + \frac{b(X_t)}{b_x(X_t)} (\exp\{b_x(X_t) \Delta t\} - 1)$$

(solve for $K_t$)

This gives

$$\mathcal{L}(X_{t+\Delta t} \mid X_t = x) = N(E_x, V_x)$$

where

$$E_x = x + \frac{b(x)}{b_x(x)} (\exp\{b_x(x) \Delta t\} - 1) \quad V_x = \sigma^2 \frac{e^{2K_x \Delta t} - 1}{2K_x}$$
Shoji-Ozaki method

\[ dX_t = b(t, X_t)dt + \sigma(X_t)dW_t \]

Use the Lamperti transform to transform this SDE into

\[ dX_t = b(t, X_t)dt + \sigma dW_t \]

Use a “better” local approximation for \( b(t, X_t) \) using first and second derivatives of \( b(\cdot, \cdot) \). The corresponding discretization is given by

\[ X_{t+\Delta t} = A(X_t)X_t + B(X_t)Z \]

where

\[
A(X_s) = 1 + \frac{b(s, X_s)}{X_s L_s} (e^{L_s \Delta s} - 1) + \frac{M_s}{X_s L_s^2} (e^{L_s \Delta s} - 1 - L_s \Delta s)
\]

\[
B(X_s) = \sigma \sqrt{\frac{e^{2L_s \Delta s} - 1}{2L_s}}
\]
Where

\[ L_s = b_x(s, X_s) \quad M_s = \frac{1}{2} b_x x(s, X_s) + b_t(s, X_s); \]

Thus

\[ \mathcal{L}(X_{t+\Delta t} \mid X_t = x) = N(A(x)x, B^2(x)) \]
Asymptotics (LLN and CLT)

- In this section we present a few results about quantities like

$$\int_0^T h(X_t) \, dt \quad \text{and} \quad \int_0^T h(X_t) \, dW_t$$

as $T \to \infty$.

- These are the continuous time versions of

$$\sum_{i=1}^n h(X_i)$$

- When appropriately normalized, we expect the “usual” limits, if the process $X_T$ is well-behaved.

- Why do we care? Many estimators (based on high-freq. data) are expressed using such quantities, thus, one can expect to establish asymptotic properties of these estimators.
Preliminaries

\[ dX_t = S(X_t)dt + \sigma(X_t)dW_t \]

Denote

\[ \tau_a = \inf\{ t \geq 0, X_t = a \} \quad \tau_{ab} = \inf\{ t \geq \tau_a : X_t = b \} \]

Definition  The process \( X_t \) is called

- recurrent if \( P(\tau_{ab} < \infty) = 1 \);
- positive recurrent if \( E(\tau_{ab}) < \infty \).
- null recurrent if \( E(\tau_{ab}) = \infty \).
Proposition. The process \( X \) is recurrent if and only if

\[
V(x) = \int_0^x \exp \left\{ -2 \int_0^y \frac{S(u)}{\sigma(u)^2} du \right\} dy \to \pm \infty
\]

as \( x \to \pm \infty \). The recurrent process is positive if and only if

\[
G = \int_{-\infty}^{\infty} \sigma(y)^{-2} \exp \left\{ 2 \int_0^y \frac{S(u)}{\sigma(u)^2} du \right\} dy < \infty
\]

The process is null recurrent if it is recurrent and

\[
G = \infty
\]

If \( \sigma \equiv 1 \) then (2) implies (1). In this case, the condition

\[
\limsup_{|x| \to \infty} xS(x) < -\frac{1}{2}
\]

is sufficient for (1) and (2).
Example

\[ dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 dW_t \quad \theta_2, \theta_3 > 0 \]

Verify that

\[ G = \int_{-\infty}^{\infty} \sigma(y)^{-2} \exp \left\{ 2 \int_{0}^{y} \frac{S(u)}{\sigma(u)^2} du \right\} dy < \infty \]

It is also true that \( V(x) \to \pm \infty \) as \( x \to \pm \infty \). (exercise)
Ergodicity

The process $X$ is ergodic if there exists an (invariant) distribution $F(\cdot)$ such that for any measurable $h(\cdot)$ such that $E(h(\xi)) < \infty$, we have the convergence

$$\frac{1}{T} \int_0^T h(X_t) dt \to E(h(\xi)) \quad \text{as} \quad T \to \infty \quad \text{a.s.}$$

Here we assume that $\xi \sim F(\cdot)$. From now on, we assume that

$(\mathcal{RP}) \quad V(x) \to \pm \infty \text{ as } x \to \pm \infty \text{ and } G < \infty.$

**Theorem.** (Law of Large Numbers) Let the conditions $(\mathcal{RP})$ be fulfilled. Then, the process $X$ is ergodic with invariant density given by

$$f(x) = \frac{1}{G\sigma(x)^2} \exp \left\{ 2 \int_0^x \frac{S(y)}{\sigma(y)^2} \, dy \right\}$$
Example

\[ dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 dW_t \quad \theta_2, \theta_3 > 0 \]

The invariant density is

\[ f(x) \propto \exp \left\{ \frac{2}{\theta_3^2} \int_0^x (\theta_1 - \theta_2 y) dy \right\} \]

\[ = \exp \left\{ \frac{2}{\theta_3^2} (\theta_1 x - \theta_2 x^2/2) \right\} \]

\[ = \mathcal{N} \left( \text{mean} = \frac{\theta_1}{\theta_2}, \text{var} = \frac{\theta_3^2}{2\theta_2} \right) \]
Assume that \( h \in M^2_T \), that is \( \int_0^T h(t, \omega) \, d\mathbb{W}_t \) is well defined.

**Theorem.** Say there exists a (non-random) function \( \varphi(T) \) such that

\[
\varphi(T)^2 \int_0^T h(t, \omega)^2 \, dt \xrightarrow{P} \rho^2 < \infty
\]

Then,

\[
\varphi(T) \int_0^T h(t, \omega) \, d\mathbb{W}_t \Rightarrow N(0, \rho^2)
\]

(see Kutoyants, p.43)
CLT for SDEs

\[ dX_t = S(X_t)dt + \sigma(X_t)dW_t \]

Assume that the \((\mathcal{RP})\) conditions hold, i.e., \((X_t)\) is positive recurrent and LLN holds

\[
\frac{1}{T} \int_0^T g(X_t)^2 \, dt \xrightarrow{P} \mathbb{E}(g(\xi)^2) \equiv \rho^2
\]

It follows that

\[
\frac{1}{\sqrt{T}} \int_0^T g(X_t) \, dW_t \Rightarrow N(0, \rho^2)
\]
More asymptotics

One can also formulate a CLT for the ordinary integral

$$\frac{1}{\sqrt{T}} \int_0^T h(X_t) dt$$

How?

$$\frac{1}{\sqrt{T}} \int_0^T h(X_t) dt = \frac{H(X_T) - H(X_0)}{\sqrt{T}} - \frac{1}{\sqrt{T}} \int_0^T H'(X_t) \sigma(X_t) d\mathbb{W}_t$$

where

$$H(x) = \int_0^x \frac{2}{\sigma(y)^2 f(y)} \int_{-\infty}^y h(v) f(v) dv dy$$
Assume that $X^T = \{ X_t, \ t \in [0, T] \}$ is a diffusion process satisfying

$$dX_t = S(\theta, X_t)dt + \sigma(\theta, X_t)dW_t \quad X(0) = X_0 \quad t \in [0, T]$$

Assume that we observe an entire function $\{ X(t) \ t \in [0, T] \}$.

**Goal:** estimate $\theta$. 
Estimating the diffusion coefficient

Given \( \{X(t) \mid t \in [0, T]\} \), \( \sigma(\theta, X_t) \) can be estimated with high accuracy, based on the quadratic variation. We have

\[
\sum_{j=0}^{n-1} \left(X(u_{j+1}) - X(u_j)\right)^2 \to \int_0^t \sigma(\theta, X_s)^2 \, ds
\]

where \( 0 = u_0 < u_1 < \cdots < u_n = t \) is a partition of \([0, t]\).

It follows that the RHS (the limit) can be “approximated” with any level of accuracy. Thus, the diffusion coefficient is determined.
Example: OU model

\[ dX_t = (\theta_1 - \theta_2 X_t) dt + \theta_3 dW_t \quad t \in [0, T] \]

The quadratic variation of an observed path \( X^T \) is

\[ \int_0^T \theta_3^2 ds = \theta_3^2 T \approx \sum_{i=1}^N (X(t_i) - X(t_{i-1}))^2 \]

where \( 0 = t_0 < t_1 < \cdots < t_N = T \) is a fine partition of \([0, T]\).

The RHS above can be evaluated with any level of precision, and thus one can estimate

\[ \hat{\theta}_3^2 = \frac{1}{T} \sum_{i=1}^N (X(t_i) - X(t_{i-1}))^2 \]
A short simulation

\[ \theta_1 = 1, \theta_2 = 2, \theta_3 = 1, \ T = 10, \]

The estimate is \[ \hat{\theta}_3^2 = \frac{1}{T} \sum (X_{i+1} - X_i)^2. \]

- 1,000 discretization steps \[ \hat{\theta}_3 = 1.033 \]
- 10,000 discretization steps \[ \hat{\theta}_3 = 1.006 \]
Estimating the drift coefficient

\[ dX_t = S(\theta, X_t)dt + \sigma(X_t)dW_t \quad X(0) = X_0 \quad t \in [0, T] \]

Assume that we observe an entire function \( X^T = \{X(t) \mid t \in [0, T]\} \).
(note that no parameters appear in the diffusion coefficient)

We work under the assumption that \( X^T \) was generated by a true model, say

\[ dX_t = S(5, X_t) + \sigma(X_t)dW_t \quad X(0) = X_0 \quad t \in [0, T] \]

How do we estimate \( \theta \)?
Likelihood function

\[ dX_t = S(\theta, X_t)dt + \sigma(X_t)dW_t \quad X(0) = X_0 \quad t \in [0, T] \]  
(3)

\[ dX_t = S(\theta^*, X_t)dt + \sigma(X_t)dW_t \quad X(0) = X_0 \quad t \in [0, T] \]  
(4)

- The observed path \( X^T = \{X_t, t \in [0, T]\} \) is an element of \( C([0, T]) \) and is generated by model (4), \( \theta^* = \text{truth} \).
- Model (3) induces a probability measure \( Q_\theta \) over \( C([0, T]) \).
- Model (4) induces a probability measure \( Q_{\theta^*} \) over \( C([0, T]) \).
- The likelihood function is

\[ \frac{dQ_\theta}{dQ_{\theta^*}}(X^T) \]
Recall the Girsanov formula

\[
\frac{dQ_\theta}{dQ_{\theta^*}}(X^T) = \exp \left\{ \int_0^T \frac{S(\theta, X_t) - S(\theta^*, X_t)}{\sigma(X_t)^2} \, dX_t \right. \\
\left. - \frac{1}{2} \int_0^T \frac{S(\theta, X_t)^2 - S(\theta^*, X_t)^2}{\sigma(X_t)^2} \, dt \right\}
\]

The Maximum Likelihood Estimate (MLE) is then

\[\hat{\theta}_{MLE} = \arg\max_{\theta} \frac{dQ_\theta}{dQ_{\theta^*}}(X^T)\]
Example: OU model

\[ dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 d\mathbb{W}_t \]
\[ dX_t = (\theta_1^* - \theta_2^* X_t)dt + \theta_3 d\mathbb{W}_t \]

Data \( X^T = \{X_t, t \in [0, T]\}. \) Here \( \theta = (\theta_1, \theta_2), \theta^* = (\theta_1^*, \theta_2^*) \) and \( \theta_3 \) is assumed known (from the quadratic variation). Maximize the likelihood function wrt \( \theta \). For this example, assume that \( \theta_2 \) is also known, \( \theta_2 = \theta_2^* \).

\[
\frac{d\mathbb{Q}_\theta}{d\mathbb{Q}_{\theta^*}}(X^T) = \exp \left\{ \int_0^T \frac{S(\theta, X_t) - S(\theta^*, X_t)}{\sigma(X_t)^2} \ dX_t - \frac{1}{2} \int_0^T \frac{S(\theta, X_t)^2 - S(\theta^*, X_t)^2}{\sigma(X_t)^2} \ dt \right\}
\]

Note: \( \sigma(X_t) = \theta_3 \) and \( S(\theta, X_t) = \theta_1 - \theta_2 X_t. \)
Example

Ignore terms in $\theta^*$, $\theta_3$, maximize

$$
\int_0^T S(\theta, X_t)dX_t - \frac{1}{2} \int_0^T S(\theta, X_t)^2 dt
$$

Ignore terms in $\theta_2$ (assumed known here), maximize

$$
\int_0^T \theta_1 dX_t - \frac{1}{2} \left( t\theta_1^2 - 2 \int_0^T \theta_1 \theta_2 X_t dt \right)
$$

Observe the quadratic, which is maximized at

$$
\hat{\theta}_{MLE1} = \frac{X_T - X_0 + \theta_2 \int_0^T X_t dt}{T}
$$
Example: asymptotics

\[ \hat{\theta}_1^{MLE} = \frac{X_T - X_0 + \theta_2 \int_0^T X_t dt}{T} \]

\[ = \frac{1}{T} \left( \int_0^T \theta^*_1 - \theta^*_2 X_t dt + \theta_3 \int_0^T dW_t + \theta_2 \int_0^T X_t dt \right) \]

\[ = \theta^*_1 + \frac{\theta_3}{T} W_T \]

And thus

\[ \sqrt{T}(\hat{\theta}_1^{MLE} - \theta^*_1) = \theta_3 \frac{W_T}{\sqrt{T}} = N(0, \theta^*_3) \]

Exercises:

- assume that \( \theta_1 \) is known and estimate \( \theta_2 \)
- assume that both \( (\theta_1, \theta_2) \) are unknown
Statistical inference based on discrete data

Consider a diffusion process

\[ dX_t = S(\theta, X_t)dt + \sigma(\theta, X_t)dW_t \quad X(0) = X_0 \quad t \in [0, T] \]

Assume that we observe

\[ X = (X(t_0), X(t_1), \ldots, X(t_n)) = (X_0, X_1, \ldots, X_n) \quad \text{where } X_i = X(t_i) \]

and \( 0 = t_0 < t_1 < \cdots < t_n = T \).

Goal: estimate \( \theta \).
Likelihood inference

Notation $[\cdot]$ — “distribution of . . .” (typically referring to the pdf)

Inference on $\theta$ is based on the likelihood function

$$L_n(\theta) = [X \mid \theta] = [X_0, X_1, \ldots, X_n \mid \theta]$$

$$= [X_0 \mid \theta] \prod_{i=1}^{n} [X_i \mid X_{i-1}, \theta] \quad \text{(Markov process)}$$

where $[X_i \mid X_{i-1}, \theta] = \text{probability density of } X_i \text{ given } X_{i-1}, \theta \text{ and } t_{i-1}, t_i$.

We use the general notation

$$p_\theta(x \mid \Delta, X_t = x_0) = P_\theta(X_{t+\Delta} \in dx \mid X_t = x_0)$$

This quantity is called the transition density, and plays an extremely important role.

$$L_n(\theta) = p_\theta(X_0) \prod_{i=1}^{n} p_\theta(X_i \mid \Delta_i, X_{i-1})$$
Maximum likelihood estimation

- The MLE of $\theta$ is defined as

$$\hat{\theta}_{MLE} = \arg\max_{\theta} L_n(\theta) \quad \text{or} \quad \hat{\theta}_{MLE} = \arg\max_{\theta} \log(L_n(\theta))$$

- Exact likelihood inference can only be done in a handful of cases, where the transition density is known.

(recall)

$$L_n(\theta) = p_{\theta}(X_0) \prod_{i=1}^{n} p_{\theta}(X_i | \Delta_i, X_{i-1})$$
OU model

\[ dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 dW_t \]
\[ X(0) = X_0, \ t \in [0, T] \ \theta_2, \theta_3 > 0 \]

The transition density is

\[ p_\theta(x \mid X_t = x_0, \Delta) = \phi(x; m_\theta(x_0, \Delta), \nu_\theta(x_0, \Delta)) \]

\( \phi(x; \mu, \gamma^2) \) is the Gaussian pdf with mean \( \mu \), variance \( \gamma^2 \) evaluated at \( x \).

Above,

\[ m_\theta(x_0, \Delta) = \frac{\theta_1}{\theta_2} + \left( x_0 - \frac{\theta_1}{\theta_2} \right) e^{-\theta_2 \Delta} \]

\[ \nu_\theta(x_0, \Delta) = \frac{\theta_2^2 (1 - e^{-2\theta_2 \Delta})}{2\theta_2} \]

- Likelihood inference will work well, without much trouble
- In certain cases, once can even get closed form expressions for \( \hat{\theta}_{MLE} \) (\( \theta_1 = 0 \))
- Asymptotic properties of \( \hat{\theta}_{MLE} \) can also be obtained, typically one requires \( n\Delta_n \to \infty \).
**GBM model**

\[ dX_t = \theta_1 X_t dt + \theta_2 X_t dW_t, \quad X(0) = X_0, \ t \in [0, T] \quad \theta_2 > 0 \]

The transition distribution in this case is

\[ p_\theta(x \mid X_t = x_0, \Delta) \sim \log - \text{Normal}(\text{mean} = m_\theta(x_0, \Delta), \text{var} = v_\theta(x_0, \Delta)) \]

where

\[ m_\theta(x_0, \Delta) = x_0 e^{\theta_1 \Delta} \]

\[ v_\theta(x_0, \Delta) = x_0^2 e^{2\theta_1 \Delta} \left( e^{2\theta_2 \Delta} - 1 \right) \]

**CIR model**

\[ dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 \sqrt{X_t} dW_t \quad X(0) = X_0, \ t \in [0, T] \quad \theta_1, \theta_2, \theta_3 > 0 \]

The transition density in this case is a non-central \( \chi^2 \).
Pseudo-likelihood methods

- In most cases, the transition density is unknown and thus the likelihood function is intractable.

- In all these cases, one has to resort to a way to approximate the likelihood function.

- These pseudo-likelihood methods are of four types:
  - methods based on numerical discretizations of the SDE
  - methods based on a simulated likelihood
  - closed form approximations (not included in these notes)
  - methods based on exact sampling (EA) algorithms
Likelihood approximations based on the Euler scheme

\[ dX_t = S(\theta, X_t)dt + \sigma(\theta, X_t)dW_t \quad X(0) = X_0 \quad t \in [0, T] \]

Consider the Euler approximation

\[ X_{t+\Delta} - X_t = S(\theta, X_t)\Delta + \sigma(\theta, X_t)(W_{t+\Delta} - W_t) \]

Observing that \( W_{t+\Delta} - W_t \) is independent of \( X_t \), it follows that

\[ X_{t+\Delta} | X_t, \theta \sim \mathcal{N}\left( X_t + S(\theta, X_t), \Delta\sigma(\theta, X_t)^2 \right) \]

The transition density is then approximated with

\[ p_\theta(x | \Delta, X_0 = x_0) \approx \phi(x; m_\theta(x_0, \Delta), v_\theta(x_0, \Delta)) \]

where

\[ m_\theta(x_0, \Delta) = x_0 + S(\theta, x_0)\Delta \quad v_\theta(x_0, \Delta) = \Delta\sigma(\theta, x_0)^2 \]

\( \phi(x; \mu, \gamma^2) \) is the Gaussian pdf with mean \( \mu \), variance \( \gamma^2 \) evaluated at \( x \).
The effect of $\Delta$

Consider the OU process

$$dX_t = (\theta_1 - \theta_2 X_t)dt + \theta_3 dW_t \quad X(0) = X_0, \ t \in [0, T] \ \theta_2, \theta_3 > 0$$

Recall that the exact transition density is Gaussian, with parameters

$$m_\theta(x_0, \Delta) = \frac{\theta_1}{\theta_2} + \left(x_0 - \frac{\theta_1}{\theta_2}\right) e^{-\theta_2 \Delta}$$

$$v_\theta(x_0, \Delta) = \frac{\theta_3^2(1 - e^{-2\theta_2 \Delta})}{2\theta_2}$$

The Euler transition density is also Gaussian, with parameters

$$m_\theta(x_0, \Delta) = x_0(1 - \theta_2 \Delta) + \theta_1 \Delta$$

$$v_\theta(x_0, \Delta) = \theta_3^2 \Delta$$

The two are “close”, only as $\Delta \to 0$. 
Simulated likelihood approximations

Focus again on the transition density and note that, informally

$$[X_{\Delta} | X_0] = \int [X_{\Delta}, X_{\Delta-\delta} | X_0] \; dX_{\Delta-\delta}$$

$$= \int [X_{\Delta} | X_{\Delta-\delta}] [X_{\Delta-\delta} | X_0] \; dX_{\Delta-\delta}$$

$$= \mathbb{E}(p_\theta(X_\Delta | X_{\Delta-\delta}, \delta))$$

where the expectation is taken wrt $X_{\Delta-\delta}$ (and $X_{\Delta}$ is held fixed).

Idea :

- If $\delta$ is very small, then $p_\theta$ can be approximated with the Euler transition density
- The expectation above can be estimated using a Monte Carlo approach (sample many many $X_{\Delta-\delta}$ . . . )
Importance sampling approach

Formally,

\[
p_\theta(x \mid x_0, \Delta) = \\
= \int p_\theta(z_1, z_2, \ldots, z_N \mid x_0, \delta)p_\theta(x \mid z_N, \delta)dz_1dz_2 \ldots dz_N \\
= \int \frac{p_\theta(z_1 \mid x_0, \delta)p_\theta(z_2 \mid z_1, \delta) \ldots p(z_N \mid z_{N-1}, \delta)p_\theta(x \mid z_N, \delta)}{q(z_1, z_2, \ldots, z_N)} q(z_1, z_2, \ldots, z_N)dz_1dz_2 \ldots dz_N
\]

where \(q(\cdot)\) is some importance sampling density. The choice of \(q(\cdot)\) is very important!

- The **main** idea is to make \(q\) depend on \(x\) and \(x_0\) !!!!
- Elerian(2001) suggests a multivariate Gaussian (or \(t\)) distribution
- Durham and Gallant (2002) propose Brownian Bridge samplers
- Stramer and Yan (2007) presents a good overview of this approach.
Approximations based on the exact sampling algorithms

EA (1,2 and 3) is a series of algorithms giving exact (no discretization error) draws from a large class of diffusion processes. See Beskos and Roberts (2005), Beskos et al. (2006), Beskos et al. (2008)

EA algorithms are rejection sampling algorithms!

Recall the likelihood function

\[ L_n(\theta) = p_{\theta}(X_0) \prod_{i=1}^{n} p_{\theta}(X_i | \Delta_i, X_{i-1}) \]

Idea: given \( X_{i-1}, X_i, \Delta, \theta \), one can simulate a random variable \( \Psi \) such that

\[ E(\Psi) = p_{\theta}(X_i | X_{i-1}, \Delta) \]
Approximations based on the exact sampling algorithms

That is, each contribution to the likelihood function can be estimated unbiasedly: simulate iid $\Psi_1, \ldots, \Psi_B$ and estimate

$$\hat{p}_\theta(X_i \mid X_{i-1}, \Delta) = \frac{1}{B} \sum_{j=1}^B \psi_j$$

This leads to an unbiased estimate of the likelihood function, which can be consequently optimized over $\theta$.

Why EA? The idea is that

$$p_\theta(X_i \mid X_{i-1}, \Delta) = \mathbb{E}("\text{acceptance probability from EA algorithm"})$$